

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



**THIS PAGE BLANK (USPTO)**



589 AsnProLysGlnProThrAsnPropheLeuGluMetIleLysPheLeuLeu 605  
 173 TTCCCTAAGCAGCGGACAAATCCATTCTAGAGATGGTGAATTCCTGCT 222  
 605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuI 622  
 223 GGAGAGAATTGGCACCAGTCACATGCTGCTGAGGCAATAAGTGCCTGG 272  
 622 leLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlu 638  
 273 TTAAGCTGTAATAAGTCAATTAAGGCACTGCTGATGATGAAGATGAG 322  
 639 GlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysVa 655  
 323 GGAGTAACCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 372  
 655 lleuSerPheThrHisProIleSerPheIleSerAlaGluThrPheGluS 672  
 373 TTATCATTTACTCAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 422  
 672 erLeuLeuAlaCysLeuLysMetAspGluLysValAlaGluAlaAla 688  
 423 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 472  
 689 LeuGlnIlePheLysAsnThrGly 696  
 473 ATACAGATCTTCAGGAACACAGCA 496

seq\_name: gb\_est73:BE306904

seq\_documentation\_block:  
 LOCUS BE306904 557 bp mRNA EST 26-OCT-2000  
 DEFINITION 601087447F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3482314 5',  
 mRNA sequence.  
 ACCESSION BE306904  
 VERSION BE306904.1 GI:9161192  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 557)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Pict: L14M8511 row: n column: 11  
 High quality sequence stop: 546.  
 Location/Qualifiers  
 1. 557  
 /organism="Mus musculus"  
 /strain="C57/B6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3482314"  
 /clone\_lib="NCI\_CGAP\_Mam5"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stages="7 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

FEATURES  
 source

seq\_name: gb\_est10:AA630616  
 seq\_documentation\_block:  
 LOCUS AA630616 391 bp mRNA EST 06-MAR-1998  
 DEFINITION acile07:sl StrataGene HeLa cell s3 937216 Homo sapiens cDNA clone  
 IMAGE:856164 3', mRNA sequence.  
 ACCESSION AA630616  
 VERSION AA630616.1 GI:2553227  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 391)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

ORIGIN

alignment\_scores:  
 Quality: 670.00 Length: 185  
 Ratio: 4.012 Gaps: 4  
 Percent Similarity: 90.270 Percent Identity: 76.216  
 alignment\_block:  
 US-09-512-581-2 x BE306904 ..  
 Align seg 1/1 to: BE306904 from: 1 to: 557

781 PheAlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLe 797  
 3 TTTGGCTCCCGCATGAATCTGTAGTGGCAAACTTTATTGTTAAAGACCT 52  
 797 uLeuMetAsnAspArgLeuProGly..LysLysThrThrLysLeuTrpVa 813  
 53 TCTAATGAACGACAGGTCAACAGGTGACGAAGAATGGAACAACTTATGGTC 102  
 813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleL 830  
 103 TCCAGATGAGGAAGTTCTCTGGAAGTTCTAGCAAGGTACAGGCAATTA 152  
 830 ysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSer 846  
 153 AACTTCTGTTAAGGTGGCTGTTGGGTATGAAAAACAACCAAGTCTAAATCT 202  
 847 GlyThrSerThrLeuArgLeuLeuThrIleLeuHisSerAspGlyAs 863  
 203 GCCAATCAACCTCTCGGTATTATTCAGCCATGCTGGTTAGTAGGGCGGA 252  
 863 pLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgL 880  
 253 CCTGACAGACCAAGAGGATCAGTAATCTGATATGCTCGCTTGGAT 302  
 880 euAlaAlaGlySerAlaIleValLysLeuAlaGlnProCysTyrHis 896  
 303 TAGCTGCTGTTAGTGCATAAATGAAGCTTGTCCAGGAACCTTGTACCAT 352  
 897 GluIleIleThrLeuGluGlnTyrGlnLeu.CysAlaLeuAlaIleAsnA 913  
 353 GAAATATTACCACCAAGACAGTTTCAGCTACTGTCACCTGGTTATTAAAG 402  
 913 spGluCysTyrGlnVal.ArgGlnValPheAlaGlnLysLeuHisLysG 929  
 403 ATGAGTGCTACCAAGTAACGGCAGATATTGGCCAGAACCTTCATAAGGC 452  
 929 yLeuSerArgLeuArgLeuProLeuGluThrMetAlaIleCysAlaLeu 946  
 453 ACTTGTGAAGTTGCTTCTCCCACTGGAGTATATGGCCATCTTTGCTTTGT 502  
 946 ysAlaLysAspProValLys.GluArgArgAlaHisAlaArgGlnCys 961  
 503 GTCCCAAGACCTCTGTGAACGACAGACAGACACGCTCGGCAGTGT 550

BASE COUNT 160 a 118 c 132 g 147 t



100

TGTAACAGGCCGAACAGTGTGTGAGGGAATC

[illegible]

13. S. HILLIER, Steptoe

seq_documentation_block:				
LOCUS	AJ558326	498 bp	mRNA	EST
DEFINITION	f78906.x1 Zebrafish Washu MPIMG EST Danio rerio CDNA 3', mRNA			
sequence.				
ACCESSION	AJ558326			
VERSION	AJ558326.1	GI:4508564		
KEYWORDS	EST.			
SOURCE	zebrafish.			
ORGANISM	Danio rerio			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Actinopterygii; Neopterygii; Rasbora; Euteleostei; Ostariophysi;			
	Cypriniformes; Cyprinidae; Rasborinae; Danio.			
REFERENCE	1 (bases 1 to 498)			
AUTHORS	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy			
	S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood			
	K., Steptoone,M., Theising,B., Allen,M., Bowers,Y., Person,B.,			
	Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,			







```

seq_documentation_block:
LOCUS      BF107038             818 bp      mRNA           19-OCT-2000
DEFINITION 601824203F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043905 5',
            mRNA sequence.
ACCESSION  BF107038
VERSION     BF107038.1 GI:10889663
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 818)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            CDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM867 row: f column: 02
            High quality sequence stop: 630.
            Location/Qualifiers
            1..818
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4043905"
               /clone_lib="NIH_MGC_79"
               /lab_host="pH108 (T1 phage-resistant)"
               /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
               Site_1: SfII (ggccattatggcc); Site_2: SfII (ggccattatggcc)
               ); 5' and 3' adaptors were used in cloning as follows: 5'
               adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
               sequence: 5'-ATTCTAGAGCGGCGGCGGCGACATG-dT(30)BN-3'
               (where B = A, C, or G and N = A, C, G, or T). Average
               insert size 1.3 kb (range 0.5-4.0 Kb). 15/15 colonies
               contained inserts by PCR. This library was enriched for
               full-length clones and was constructed by Clontech
               Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
               Library."
BASE COUNT  251 a 173 c 185 g 209 t
ORIGIN
alignment_scores:
Quality: 678.50 Length: 274
Ratio: 3.043 Gaps: 9
Percent Similarity: 81.387 Percent Identity: 61.679
alignment_block:
US-09-512-581-2 x BF107038 ..
Align seg 1/1 to: BF107038 from: 1 to: 818
460 AlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCy 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 GCTCAGTATCTGTGCCCAACCTGGAAACAGACAGAGAGAGAAATG 68
476 sLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuA 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
69 CTTATATTACTTATCTGTTGGATCCAAATGCTGTAAGACTCTCA 118
493 snGluMetTrpLysCysGlnAsnLeuArgHisGlnValLysAspLeu 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
119 ACGAAATGTGGAAGTGTACAGACATGCTTCGACCATGTACGCCAA... 165
510 Leu.AspleuLeuLysGlnProTysThrAspAlaSerValLysAlaLeuP 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

166 CTATGATTGCACAAGCAGCGCTACATCAGAGCGCTAACTGTCTGCCATGT 214
526 heserLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
215 T.GGAAACATGATGACCATAGCAAGAAATTTGCCCTGACCCCGGAAAGCA 264
543 GlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIle 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
265 CAAGA.TTTCTGGAAGAATTTAACCCAGGTCTTCGCCGCGATGATGAGAACT 313
559 eaArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnA 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 TCGGTCTCAGTTGGAGTATTAAATTAGCCCAACCTGTTCTTGTCAACAAG 363
576 laGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGln 592
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 CAGATATTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTTAAGCAA 413
593 ProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluAfrGileAl 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 CCAACAATCCTTTCTAGAGATGCTCAATTTCTGTG.GAAACAATCGC 462
609 aProValHisIleAspThrGluSerIleSerAlaLeuLys.GlnVal 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 ACCTGTGCACATTGATTGAGAAAGCCATAAGTGCACTAGTGAATTTGATG 512
626 AsnLysSerIleAspGlyThr.AlaAspAspGluAspGluGlyValProT 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
513 AATAAGTCAATAGAGGAGCAGGACAGATGATGAAGAGGAGGTGTAACGT 562
642 hrAspGlnAlaIleArgAlaGlyLeuLeuLeuLysValLeuSerPhe 658
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
563 CCAGATACAGATATCCGTTAGAGATGCTCAATTTCTTAAGTTCTGTC.TTT 611
659 ThrHisProLysSerPheHisSerAlaGluThrPheGluSerLeuLeuAl 675
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
612 ACACATCCTACCTCGTTC...ACTCTGCAGAGACTATGAGTCTCTTACA 658
675 acYsLeuLysMetAspAspGluLysValAlaGlu.....AlaAlaLeuG 690
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
659 GTGCCTAAAG.....GGGTGCAGGTGGCGAAGTGTATATCCACTTTTA 702
690 InIlePheLysAsn...ThrGlySerLysIleGluGluAspPheProHisI 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
703 CGATTCGCTCCAAATGACGACTTTCGATAGATGACTTATTCCTTTC 752
706 leArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyPro 722
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
753 TAAAGCA.....AGGGGTCCC 769
723 ProArgGlnAlaLys 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
770 CCACCACAGGGGCGT 784
seq_name: gb_gss34:AZ902024
seq_documentation_block:
LOCUS      AZ902024             751 bp      DNA           05-MAR-2001
DEFINITION RPCI-24-158H23-TJ RPCI-24 Mus musculus genomic clone RPCI-24-158H23
            , DNA sequence.
ACCESSION  AZ902024
VERSION     AZ902024.1 GI:13220969
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 751)
AUTHORS     Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akınret,B., Levis,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebreyegorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)

```



## alignment\_block:

US-09-512-581-2 x AA939593

Align seg 1/1 to: AA939593 from: 1 to: 486

1143 LysGlnSerGlnThrLysSerArgMetGluThrValSerAsnAlaSe 1159  
 2 AAACAGTCTCAGACCAAAATCATCAAGAATGGAACTGTGAGCAACGCAAG 51  
 1159 rSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspS 1176  
 52 CAGCAGCTCCACCCCAAGCTCTCTGGAAGGATCAAGGGAGGCTTGATA 101  
 1176 exSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerPro 1192  
 102 GCTCTGAATGGATCAGATGAGAAATGAAGATTATACAAATGCTTCACCT 151  
 1193 LeuProGlyLysLysSerAspLysArgAspSerAspLeuValArgSe 1209  
 152 TTGCCAGGAAAAAAGTCACAGAGAGAGAGCCCTGATCTT.....TC 195  
 1209 rGluLeuGluLysProArgGlyArgLysLysThrProValThrGluGlnG 1226  
 196 TGAGTTGGAGAGACCTAGAGTCGGAAAAAGACCTGTACAGACCCTG 245  
 1226 luGluLysLeuGlyMetAspLysThrLysLeuValGlnGluGlnLys 1242  
 246 AAGAGAAATTAGTATGATGATGACCTAACTAAGTTGGTACAGGAACAGAAA 295  
 1243 ProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSe 1259  
 296 CCTAAAGGAGCAGTCAGGAGCGGAAAGAGGCGGTACAGCTCAGACTC 345  
 1259 rAspGluGlnGlnTrpProGluLysArgLeuLysGluAspIleLeuG 1276  
 346 AGACGAGCAGCAGTGCCCTGAGGAGAGAGGACCAAGAGGAGCTCTCG 395  
 1276 luAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArg 1292  
 396 AAAATGAGGATGAGCAGAGACAGCCCAACCAAGGCAAGAGAGAGCAGG 445  
 1293 ProProLysProLeuGlyGlyLysThrProLysGlu 1304  
 446 CA.CCAAAACCTCTTGGTGGGGGAGCAGCAAGGAAG 480

seq\_name: gb\_est88:BF451878

seq\_documentation\_block:

LOCUS BF451878 580 bp mRNA EST 29-DEC-2000  
 DEFINITION uz80904.y1 NCI-CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3675414 5' similar to TR:Q9Y451 Q9Y451 ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN. ; mRNA sequence.  
 ACCESSION BF451878  
 VERSION BF451878.1 GI:11518047  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 580)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov/image/html/lresources.shtml](http://image.llnl.gov/image/html/lresources.shtml)

MGI:1436182

Seq primer: -40RP from Gibco

High quality sequence stop: 376.

## FEATURES

## source

1..580  
 /organism="Mus musculus"  
 /strain="C57BL/6J (f1er1)"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3675414"  
 /clone\_lib="NCI-CGAP\_Lu29"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 /stem\_cell\_origin="Stem cell origin."  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 161 a 111 c 161 g 147 t  
 ORIGIN

## alignment\_scores:

Quality: 680.50 Length: 159  
 Ratio: 4.629 Gaps: 1  
 Percent Similarity: 92.453 Percent Identity: 83.648

## alignment\_block:

US-09-512-581-2 x BF451878

Align seg 1/1 to: BF451878 from: 1 to: 580

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrThrProPr 17  
 104 ATGGCTCATTCACAGACAGACCAACGATGGGAAATTAATTACCTCC 153  
 17 OGlyValLysGluIleSerAspLysLysSerLysGluGluMetValArg 34  
 154 TGGAGTCAAGGAAATCTCAGATAAAATCTCTAAAGAGGAGATGGTGAGC 203  
 34 rGluLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50  
 204 GGTAAAGATGGTTGTAAAACTTTCATGACATGGACATGGACGACTCTGAA 253  
 51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67  
 254 GAGGAAAGGAACTTATCTAAACCTAGCTTTACATCTTCTTCTCCTT 303  
 67 ePheLeuLysHisProGlyLysAspValArgLeuValAlaCysCysL 84  
 304 CTCTCTCAAGCATCTCTGATAAAGATGTTCTGTTACTGGTGGCTTGTG 353  
 84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100  
 354 TTGCTGATATTTTCAGGATTTATGCTCTGAGGCTCCTTACAGCTCTGCA 403  
 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117  
 404 GATAGACTTAAGGATATATGTTATTAACAAGGCGACTGAAGGACT 453  
 117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu 134  
 454 AGAAGATACACAGAGCCCTCGATTTAAATTTAGGATTTGTATGATGAG 503  
 134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150  
 504 GCATGGCGTGGGTTAAATCATATAACATGCTTGGAGTTGGAGAGCGC 553  
 151 AsnGlu.....IlePheThrGln 156  
 554 ATGGAATGCTTACTCATCTTTTACAGAG 580

seq\_name: gb\_est83:BF107038



**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 768 Std Error: 0.00  
Seq primer: -40ml3 fwd ET from Amersham  
High quality sequence stop: 411.

# FEATURES

**Source**  
1. .483  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1525851"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19w, testis NHT, and B-cell  
NCI-CGAP-GCBI) were mixed, and ss circles were made in  
vitro. Following MAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
**BASE COUNT** 149 a 101 c 78 g 155 t  
**ORIGIN**

# alignment\_scores:

Quality: 717.00 Length: 158  
Ratio: 4.717 Gaps: 0  
Percent Similarity: 96.203 Percent Identity: 84.810

# alignment\_block:

US-09-512-581-2 x AA913927/rev ..

Align seg 1/1 to reverse of: AA913927 from: 1 to: 483

250 AspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrrAsnIleAs 266  
|||||  
483 GATTGTGCAGACATGTTATGATCTGATTCAGGAACCTTTTGGCTATAGA 434  
266 pSerHisLeuLeuSerValLeuProGlnLeuGluPheLysLeuLys 283  
|||||  
433 TCCTCATTTATTATTCGCTCATCGCCACACAGCTGAATTCAAACTAAGA 384  
283 erAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMet 299  
|||||  
383 GCAATGATGAGAGAGAGCTAGCTTCTTCGACTTCAGCTTAATG 334  
300 PheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTpgL 316  
|||||  
333 TTTCGCTCCAAAGATCTGATTTGGCAGACACAGAACTCTCTTTGGCA 284  
316 nCysTyTLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluC 333  
|||||  
283 ATGTTTCTTGGCAGATTAAATGATATTCATGTTCTCTGAGATTAGAAA 234  
333 yValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 349  
233 GTGTGAATTTGCCAGTCATTGTTAATGAATCACCAGATTTAGCGAAG 184  
350 AspLeuThrGluTyTLeuLysValArgSerHisAspProGluGluAlaI 366  
183 GATCTCACAGAAATTTAAAGGTTAGATCATGATCCAGCAAGAGCTAT 134  
366 eArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleL 383  
|||||

133 TCGTCATGATGTCATCTGTTACTATAATAACAGCTGCCAAGAGGACCTGG 84  
383 euleuValAsnAspHisLeuAsnPhelValArgGluArgThrLeuAsp 399  
|||||  
83 CCTTAGTAATGATGACAGCTGCTGGCTTTGTAGGGAAGAAACACTGGAT 34  
400 LysArgTrpArgValArgLysGlu 407  
|||||  
33 AACCGTGGCGAGTAAGAAAAA 10

seq\_name: gb\_est13:AA939593

seq\_documentation\_block: 486 bp mRNA EST 01-MAY-1998  
LOCUS AA939593  
DEFINITION Y51F06.r1 Soares\_thymus\_2NbMT Mus musculus cDNA clone  
IMAGE:1330019 5', mRNA sequence.

ACCESSION AA939593  
VERSION AA939593.1 GI:3100370  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 486)

# AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

# TITLE

The WashU-HHMI Mouse EST Project

# JOURNAL

Unpublished (1996)  
Contact: Marra M/Mouse EST Project

# COMMENT

WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 485.

# FEATURES

## Source

1. .486  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1330019"  
/clone\_lib="Soares\_thymus\_2NbMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAGTGGAGCGCCGCTTTTGTGTGTGTGTGTGTGTGTGT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through two  
rounds of normalization, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

**BASE COUNT** 179 a 103 c 133 g 71 t  
**ORIGIN**

# alignment\_scores:

Quality: 717.00 Length: 162  
Ratio: 4.626 Gaps: 1  
Percent Similarity: 95.679 Percent Identity: 88.889



393 AAGAGCCCTCAATTAAATAGTATTTTATTACTTGAGAACATTGCAAT 344  
 137 rpVallysSerTyAsnIleCysPheGluLeuGluAspSerAsnGluIle 153  
 343 GGGTTAAATCATATACATATGCTTTGAGTTAGAGATAGCAATGAATC 294  
 154 PheThrGlnLeuTyArgThrLeuPheSerValIleAsnAsnGlyHisAs 170  
 293 TTTACTCAACTTTTACAGAACATATTCTCAGTTATTAACAATGGCCACAA 244  
 170 nGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysG 187  
 243 TCAGAAAGTTTATATGACATATGCTGGACCTCATGACCTCTATCATTTG 194  
 187 luGlyAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeu 203  
 193 AAGGTGATACCGTATCTCAGGAGCTCTTAGATACAGCTTTTAGTAAATCTG 144  
 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyAspLeuAlaLysAl 220  
 143 GTACCTGCCCATGAAGAACTTAAACAAGCAAGCATATGATTGGCAAGGC 94  
 220 aLeuLeuLysAsnThrAlaGlnAlaIleGluProTyIleThrPheP 237  
 93 TTTGCTCAAGAGACTGCTCAAGCTATTGACCATATATATACCAATTTT 44  
 237 heAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249  
 43 TTAACCAAGCTTCTGATGCTTGGGAAAAAAATTCGAGC 6

seq\_name: gb\_est72:BE252186

seq\_documentation\_block: 546 bp mRNA EST 13-JUL-2000  
 LOCUS BE252186 601113791F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3354263 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE252186  
 VERSION BE252186.1 GI:9123222  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 546)  
 NIH-MGC http://mhc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaaps-r@mail.nih.gov  
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LUCM156 row: f column: 24  
 High quality sequence stop: 535.  
 Location/Qualifiers  
 1..546  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3354263"  
 /clone\_lib="NIH\_MGC\_16"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

FEATURES  
 source

BASE COUNT 181 a 100 c 127 g 138 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 726.00 Length: 181  
 Ratio: 4.197 Gaps: 0  
 Percent Similarity: 95.580 Percent Identity: 76.243  
 alignment\_block:  
 US-09-512-581-2 x BE252186 ..  
 Align seg 1/1 to: BE252186 from: 1 to: 546  
 358 ArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIl 374  
 3 AGATCATCATGATCCAGAGAAGCTATTTCGTCATGATGTCATTGTTACTAT 52  
 374 eValThrAlaIleLysLysAspIleLeuValAsnAspHisLeuLeuA 391  
 53 AATAACAGCTGCCAGAGGGACCTGGCCTTAGTAAATGATCAGCTGCTTG 102  
 391 snPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGlu 407  
 103 GCCTTGAAGGGAAGAACACCTGGATAAAGCGTGGCGAGTAAGAAAGAA 152  
 408 AlaMetMetGlyLeuAlaGlnIleTyLysLysTyrAlaLeuGlnSerAl 424  
 153 GCTATGATGGTCTGGCTCAGCTTTATAGAAATACTGCTTCATGGTGA 202  
 424 aLaGlyLysAspAlaIleLysGlnIleAlaTrpIleLysAspLysLeuL 441  
 203 AGCAGGAAAGGAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAACTTC 252  
 441 euHisIleTyTyArgGlnAsnSerIleAspAspArgLeuLeuValGluArg 457  
 253 TGCATATTTATTATCAGAACAGCATTTGACACAACTGTTGGTAGAGAA 302  
 458 IlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMe 474  
 303 ATCTTTGCTCAGTATCTTCTCCCAACACCTGGAAACAGAGAGAAT 352  
 474 TyLysCysLeuTyTyArgGlnAlaThrLeuAspLeuAsnAlaValLysA 491  
 353 GAATGCTTATATATCTATATGCTAGTTGGATCCAAATGCTGTAAAG 402  
 491 laLeuAsnGluMetTyrLysCysGlnAsnLeuArgHisGlnValLys 507  
 403 CTCTCAACCAATGTGCAAGTGTCAAGACATGCTTCGGAGCCCATGACGC 452  
 508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAl 524  
 453 GAACATTATGGATTGTCACAAAGCAGCCTACATCAGAGGCTAACTGTTCGC 502  
 524 allePheSerLysValMetValIleThrArgAsnLeuProAsp 538  
 503 CATGTT.GGAAACTGATGACCATAGCAAGAAATTTTCCTGCTGAA 544  
 seq\_name: gb\_est13:AA913927

seq\_documentation\_block:  
 LOCUS AA913927 483 bp mRNA EST 10-JUN-1998  
 DEFINITION O139f02.s1 Soares\_NFL\_T-GBC\_S1 Homo sapiens cDNA clone  
 IMAGE:1525851 3', mRNA sequence.  
 ACCESSION AA913927  
 VERSION AA913927.1 GI:3053319  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 483)  
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.



/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.45 kb. Life Technologies catalog #:  
11541-018"

BASE COUNT 50 a 100 c 90 g 183 t  
ORIGIN

alignment\_scores:  
Quality: 731.00 Length: 140  
Ratio: 5.221 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-512-581-2 x AI680124/rev ..

Align seg 1/1 to reverse of: AI680124 from: 1 to: 423

```
1247 GlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnG1 1263
1248 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
1249 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
1263 nTTPProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspG 1280
1264 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
372 GTGGCCTGAGGAAAGAGAGCTCAAGAGAGATATATTAGAAAATGAAGATG 323
1280 LuGlnAsnSerProProLysLysGlyLysArgGlyArgProProLysPro 1296
1281 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
322 AACAGAAATAGTCGCCCAAAAGGGTAAAGAGAGCCAGCCACCAAAACCT 273
1297 LeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLy 1313
1298 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
272 CTGGTGGAGGTACACCAAGAGAGAGCCATACGGCTTCAGATCTGATGAC 223
1313 sGlySerLysLysLysSerGlyProProAlaProGluGluGluGluG 1330
1314 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
222 AGGAAGCAAAAGAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
1330 LuGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHis 1346
1331 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
172 AAGAAAGACAAAGTGGAAATACGGAACAGAGAGTCCAAAGCAACAGCAC 123
1347 ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAl 1363
1348 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
122 CGAGTGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 73
1363 alleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerL 1380
1364 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
72 AATTGAATCCACAGAGTCCACACACAGAGAGAGAGAGAGAGAGAGAGAG 23
1380 ysthrProSerProSerGln 1386
1381 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 373
22 AAAGCCATCACCATCACAA 3
```

seq\_name: gb\_est81:BE948523

seq\_documentation\_block:  
LOCUS BE948523 445 bp mRNA EST 03-OCT-2000  
DEFINITION UI-M-BH3-awr-g-03-0-UI-s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-awr-g-03-0-UI 3', mRNA sequence.  
ACCESSION BE948523  
VERSION BE948523.1 GI:10526282  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 445)  
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)

# MEDLINE COMMENT

97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Oligo-dT track not found, Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

# FEATURES source

Location/Qualifiers  
1..445  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-awr-g-03-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Ronaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_SEQ=None found"

BASE COUNT 134 a 83 c 74 g 154 t  
ORIGIN

alignment\_scores:  
Quality: 731.00 Length: 146  
Ratio: 5.076 Gaps: 0  
Percent Similarity: 98.630 Percent Identity: 97.260  
alignment\_block:  
US-09-512-581-2 x BE948523/rev ..  
Align seg 1/1 to reverse of: BE948523 from: 1 to: 445  
104 LysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120  
105 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 374  
443 AAGCATATATTTATGTTTAAACAGGCACTAAGGCACTAGAGATAC 394  
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAla 137  
121 CAGGAAAGTGGAAAGAGAGCCATACGGCTTCAGATCTGATGACAGCA 374



```

DEFINITION BP250017A10D2 Soares normalized bovine placenta Bos taurus cDNA
clone BP250017A10D2 5', mRNA sequence.
ACCESSION BF043498
VERSION BF043498.1 GI:10760553
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 595)
AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinaz,J., Liu,L. and Larson
,J.H.
TITLE Bovine ESTs
JOURNAL Unpublished (2000)
COMMENT Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimmi g:
Cross match from Washington University Genome Center PHRAP suite.
This sequence is vector free and at least 200 bp in length.
PCR primers
FORWARD: TATACGACTCACTATAGGG
BACKWARD: ATTAACCCCTCACTAAAG
Insert Length: 595 Std Error: 0.00
Plate: BP250017A10 row: D column: 2
Seq primer: AGCGATAACAATTTTCACACGGA
High quality sequence stop: 595.
FEATURES             Location/Qualifiers
     source
       1..595
         /organism="Bos taurus"
         /db_xref="taxon:9913"
         /clone="BP250017A10D2"
         /clone_lib="Soares normalized bovine placenta"
         /sex="female"
         /lab_host="DH10B"
         /note="Organ: placenta; Vector: pT73pac; Site_1: EcoRI;
         Site_2: NotI; The cDNA library was contributed by the
         Soares laboratory and it was constructed and normalized
         as described by Bonaldo, M.F., Lennon, G. and Soares,
         M.B. (1996), Genome Research 6(9): 791-806."
BASE COUNT 249 a 110 c 151 g 85 t
ORIGIN
1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPr 1221
|||||
2 GATCTTGTAAGTCTGAATGGCAACCAAGAGCCGCAAGAAAGACACC 51
1221 oValThrGluGlnGluGluLysLeuGlyMetAspLeuThrLysLeuV 1238
|||||
52 TGTAACAGATTCAGAGAAATTTAGGCATGGATGACCTTTCCAAAGTGG 101
1238 aLcLcGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHis 1254
|||||

```

```

102 TACAGGAACAGAAACCTAAAGCAGCTACCGTGGAGCAAGAAAGAGAGGCCCAT 151
1255 ThrAlaSerGluSerAspGluGlnGlnTPrProGluGluLysArgLeuLy 1271
|||||
152 ACAGCTTCAGATCAGCAACAGCAGCTGCGCTGAGGAAAGAGGCTCAA 201
1271 sGluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysG 1288
|||||
202 AGAAGATATACTAGAAAATGAGGATGACAGACACAGCTCCACCAAAAAAG 251
1288 lYlYsArgGlyArgProProLysProLeuGlyGlyThrProLysGlu 1304
|||||
252 GTAAAGGAGCGCCACCAACCTCTTGGTGGGGGTACCCCAAGAA 301
1305 GluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyPr 1321
|||||
302 GAGCCAACGATCAAAACATCCAAAAAGGAGCAAAAAAATCTCGACC 351
1321 oProAlaProGluGluGluGluGluGluGlnSerGlnSerGlyAsnThrG 1338
|||||
352 TTCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
1338 luGlnLysSerLysLysGlnHisArgValSerArgArgAlaGlnGln 1354
|||||
402 AGCAGAAAGTCAAAAGTAACAGCAGCAGGACTTCAAGGAGAGAGAGAGAG 451
1355 ...ArgAlaGluSerProGluSerSerAlaIleGluSerThrGlnSerTh 1370
|||||
452 AGCCGTGTAGGAGCGCTCTAAACAGCAGCTACTAAAGAAAATGACTCAAG 501
1370 rProGln 1372
502 TGAAGAA 508
seq_name: gb_est23:AI680124
seq_documentation_block:
LOCUS AI680124 423 bp mRNA EST 15-DEC-1999
DEFINITION tw65b01.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2264521 3',
mRNA sequence.
ACCESSION AI680124
VERSION AI680124.1 GI:4890306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 423)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1158 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source
       1..423
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:2264521"
         /clone_lib="NCI_CGAP_Ut3"
         /tissue_type="poorly-differentiated endometrial
         adenocarcinoma, 2 pooled tumors"

```



```
1010 AspIleGluInLeuLysAspValLysGluCysLeuTrpPheValLeuG1 1026
|||||
3 GATATTGAACAACCTAGACAGCTTAAGAGTGTCTTTGGTTGTTCTGGA 52
|||||
1026 uileLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysM 1043
|||||
53 AATACTAATGCTAAAAATGAAATAATAGTCATGCGCTTTATCAGAAGA 102
|||||
1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlnProAspAla 1059
|||||
103 TGTGTAGAAAATATTAACAACAAGAGTCTCAAGGACGAGATGATGCA 152
|||||
1060 LysMetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleI 1076
|||||
153 AAAATGAATGAAAACCTGTATCTATGTATGTATGCTTCCCATGATATCAT 202
|||||
1076 eMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValL 1093
|||||
203 CATGTCAAAAAGACCAACATATAGTTTGGAAATCTCCTAAAGACCCGGTAC 252
|||||
1093 euProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLys 1109
|||||
253 TGCAGCTCTCTTTTTCACCCAGCCTCACAAGAACTTCAGTAACACCAAA 302
|||||
1110 AsnTyrLeuProGluMetLysSerPhePheThrProGlyLysProLy 1126
|||||
303 AATATCTGCTCCAGAAATGAATATCTTTTCACTCTCTGGAAACCCAA 352
|||||
1126 sThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyL 1143
|||||
353 AACAAACCAATGTTCTAGGAGCTGTTAATAAGCCGCTTTCATCAGCAGGCA 402
|||||
1143 ysGlnSerGlnThrLysSerSerArgMetGlu 1153
|||||
403 AACAATCTCAGACAAATATCATCAGAAATGGAA 434
```

seq\_name: gb\_est8:AA511220

```
seq_documentation_block: 437 bp mRNA EST 08-JUL-1997
LOCUS AH511220
DEFINITION v556b10.r1 Soares_mammary_gland_NBMGM Mus musculus cDNA clone
IMAGE:890971 5', mRNA sequence.
ACCESSION AA511220
VERSION AA511220.1 GI:249074
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 437)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:518931
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 407.
Location/Qualifiers
i. .437
/organism="Mus musculus"
```

```
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:890971"
/clone_lib="Soares_mammary_gland_NBMGM"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/Note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCAATGGTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
```

BASE COUNT 117 a 96 c 111 g 113 t  
ORIGIN

alignment\_scores:  
Quality: 741.00 Length: 145  
Ratio: 5.110 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.310

alignment\_block:  
US-09-512-581-2 x AA511220 ..  
Align seg 1/1 to: AA511220 from: 1 to: 437

```
857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysLysSerLysPr 873
|||||
3 ATACTGCATAGTGTATGGGATTTGACAGAACAAAGGAAATTTAGTAACC 52
|||||
873 oAspMetSerArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
|||||
53 AGATATGTCAGCCCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTG 102
|||||
890 laGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeu 906
|||||
103 CACAGGAGCCCTGTTACACAGAGATCATACACTGGAGCAGTACCAGCTG 152
|||||
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
|||||
153 TGTGCATATGCCATCAATGATGAGTGTATCAAGTCAGGAGGTGTTGCG 202
|||||
923 aGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrM 940
|||||
203 TCAGAAACTTCACAAAGGCCCTTTCCCGCTTACGGCTTCCCTCTGAGTACA 252
|||||
940 eAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
|||||
253 TGGCCATCTGTGCTCTTTGTGCCAAAGACCCCTGTGAAAGAGAGCGAGCC 302
|||||
957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLe 973
|||||
303 CATGCTAGACAGTGTGCTGGTGAAGAACATCATCTGTGAGGAGGAGTACCT 352
|||||
973 uLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluT 990
|||||
353 GAACACACATGCAGCTGTTAGTGAATAATATTGCTCTCTTCTACCAAGT 402
|||||
```

seq\_name: gb\_est8:BF043498

seq\_documentation\_block:  
LOCUS BF043498 595 bp mRNA EST 10-OCT-2000







ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE917453  
BE917453.1 GI:10419093  
EST  
house mouse.  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 944)

NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM9134 row: i column: 14  
High quality sequence stop: 661.  
Location/Qualifiers

FEATURES  
source

1..944  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3964501"  
/clone\_lib="NCI\_CGAP\_Maml"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="3 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 290 a 194 c 246 g 214 t  
ORIGIN

alignment\_scores:

Quality: 751.00 Length: 252  
Ratio: 3.308 Gaps: 2  
Percent Similarity: 90.079 Percent Identity: 66.270

alignment\_block:

us-09-512-581-2 x BE917453

Align seq 1/1 to: BE917453 from: 1 to: 944

493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509  
|||||  
2 AATGAATGTGGAATGTCAGACATGCTCGAAGTCATGTGCGTGAAC 51  
|||||  
509 uLeuAspLeuLeuGlnProLysThrAspAlaSerValLysAlaIleP 526  
|||||  
52 ATG.GACTTACACAGCAGCCTACATCAGAGCGAAGTGTCTGCCATGT 99  
|||||  
526 heSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542  
|||||  
100 T.GGGAACTGATGACCATACCAAGAAATTTGCTGACCTGGAAAAGCA 149  
|||||  
543 GlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysI 559  
|||||  
150 CAAGA.TTTCTAAGAAATTTAACCAGGTCTTGGTGATGATGAGAACT 198  
|||||  
559 eArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGln 576  
|||||  
199 GAGGTCTCACTTGAATATTATATCAGCCCAACCTGTTCATGCAAGCAG 248  
|||||  
576 laGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGln 592  
|||||  
249 CTGACGT.TGTGTGAGGAAATAGCTCGGAAACTTGCAGAACTCTAAGCAG 297  
|||||

593 ProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAl 609  
|||||  
298 CCAACCAACCCCTTTCTAGACATGGTCAAAATTTCT.TTGGAAAGAAATGC 346  
|||||  
609 aProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVal 626  
|||||  
347 TCCTGTGCACATCGATTCAAGCCATAGTCAAGTGGTAAACTGATGA 396  
|||||  
626 snLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThr 642  
|||||  
397 ATAAGTCAATCGAAGGACGCGCGATGATGAAGAGAGGGGTGTCAAGTCCA 446  
|||||  
643 AspGlnAlaIleArgAlaGlyLeuGluLeuLysValLeuSerPheTh 659  
|||||  
447 GACTCAACCCCTCGCTCAGGACTTGAGCTTCTTAAGGTCTGTCTTTTCC 496  
|||||  
659 rHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAla 676  
|||||  
497 ACATCTACCTCGTTCACCTGCAGAGACATATAGTCTCTGTTCAGT 546  
|||||  
676 ystLeuLysMetAspAspGluLysValAlaGluAlaLeuGlnIlePhe 692  
|||||  
547 GCCTAAGATGGAGGATGACAAGGTAGCAGAGCTGCGATACAAATTTT 596  
|||||  
693 LysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAl 709  
|||||  
597 AGAACACAGGCGACAAATAGAAACTGAGCTTCCCGCATACGCTCCAC 646  
|||||  
709 aLeuLeuProValLeuHisLysSerLysLysGlyProArgGlnA 726  
|||||  
647 TTGATCCCCA...TTTATCATCAGACCCCAAGGGGGGACTCACAAACCAAG 693  
|||||  
726 laLysThrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThr 742  
|||||  
694 CAAGAAGTGT...CACTGGATTTCATGCATCTTCTCAAACAGGAGGTC 740  
|||||  
743 GlnPhe 744  
|||||  
741 AGGTGG 746

seq\_name: gb\_est98:BG258462

seq\_documentation\_block:

LOCUS BG258462 777 bp mRNA EST 13-FEB-2001  
DEFINITION 602379941F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4510378 5',  
mRNA sequence.

ACCESSION BG258462

VERSION BG258462.1 GI:12768191

KEYWORDS EST

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 777)

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

High quality sequence start: 4

High quality sequence stop: 678.

Location/Qualifiers

1..777

/organism="Homo sapiens"

FEATURES

source



778 oAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValL 795  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 202 AGACAGATTGCTGCTCGCTCAAAATCTTAGTGCCCAATTCATCGTTA 251  
 795 yAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 252 AAGACCTGCTAATGAACACCGGATTCCTGGCAAGAA .ACCACCAAGCTG 300  
 812 TrpValProAspGluGluValSerProGluThrMetValLysIleGlnAl 828  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 301 TGGGTCCAGATGACGAAGTCTCTCTGAAACATCAGCCCAAGATTCAGG 350  
 828 aLeLysMetMetValArgTrpLeuLeuGlyMetLysAsnHisSerL 845  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 351 TCTGAAACTGATGTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400  
 845 ySerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAsp 861  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 401 AATCTGGAACACTCAACTGCGGATGTTGACGGCAATCTGAGCAGCGAT 450  
 862 GlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgL 878  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 451 GGAGATCTGACGGACGACGAGGATGGGAGCGGACATGCTCTGCTCT 500  
 878 uArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCyst 895  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 501 GCGTCTGCGACGGCTGCTGCGATCTCTGCTGCTGCTGCTGCTGCTGCT 550  
 895 yHisGluIleIleThrLeu 901  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 551 ACCATGAGATCATCATCCCTG 570

seq\_name: gb\_est98:BG258248

seq\_documentation\_block: 896 bp mRNA EST 13-FEB-2001  
 LOCUS BG258248 602379670F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4510600 5',  
 DEFINITION mRNA sequence.

ACCESSION BG258248

VERSION BG258248.1 GI:12768064

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM10392 row: k column: 17

High quality sequence stop: 659.

FEATURES

source

1. .896

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4510600"

/clone\_lib="NIH\_MGC\_92"

/tissue\_type="embryonal carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pCMV-Sport6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

BASE COUNT 295 a 172 c 208 g 221 t

ORIGIN

alignment\_scores:

Quality: 767.00 Length: 214

Ratio: 4.037 Gaps: 1

Percent Similarity: 88.785 Percent Identity: 70.093

alignment\_block:

US-09-512-581-2 x BG258248 ..

Align seg 1/1 to: BG258248 from: 1 to: 896

605 LeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLe 621  
 :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 13 GTGGAAAGAAATCGCACCTGTGCACCATGATTCAGAAAGCCATTAAGTGCAC 62  
 621 uLeLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspG 638  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 63 AGTGAATTTGATGAATAGTCAATAGAGGGGACAGCATGATGAAGAGG 112  
 638 luGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLys 654  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 113 AGGGTGAAGTCCAGATACAGCTATCCGTTTCAGGACTTGAACCTCTTAAG 162  
 655 ValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheG 671  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 163 GTTCTGCTTTTACACATCCTACCTCGTTCACCTCTGCAGACACATATGA 212  
 671 uSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAlaA 688  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 213 GTCTTGTTTACAGTCCTTAAGAATGGAGGATGACAAAGGTAGCAGAAGCTG 262  
 688 laLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePro 704  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 263 CTATTCAAAATTTTGAATAACAGGTGCACAAATAAGAAACAGACCTTCCC 312  
 705 HisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysG 721  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 313 CAGATACGATCGACCTTAATTCCTCATTTTACATCAAAAGCAAGAGGGG 362  
 721 yProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheS 738  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 363 TACTCCACACCAAGCAAGCAAGGCTGTGCACGTGTATACACGCCATATTC 412  
 738 erSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSer 754  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 413 CAATAAAGAAAGTCCAGCTTGCACAGATTTTGTAGCCACTCAGTAGGAGT 462  
 755 LeuAspProSerAsnLeuGluHisIleThrProLeuValThrIleG 771  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 463 CTGAATGCTGATGTGCCAGACAACACTTATACTCCATTAGTTTCATTGGG 512  
 771 yHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerT 788  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 513 CCACATTTCTATGTTAGCACCAGATCAGTTTGTCTTCCCAATGAATCTG 562  
 788 rpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuPro 804  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 563 TAGTAGCAAAATTTTATTGTGAAGATCTGCTAATGAATGACAGGTCACAA 612  
 805 GlyLys..LysThrThrLysLeuTrpValProAspGluGlu 817  
 1:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 613 GGTGAACAAGATGCAACCACTGTGCTCTCCAGATGAAGAG 653

seq\_name: gb\_est80:BE917453

seq\_documentation\_block:

LOCUS BE917453 944 bp mRNA EST

DEFINITION 601664353F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3964501 5',

mRNA sequence.

29-SEP-2000



JOURNAL  
COMMENT

Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [zbrafish@watson.wustl.edu](mailto:zbrafish@watson.wustl.edu)  
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
ReSource/Zentrum für Molekulare Genetik Berlin, Germany (web address:  
[www.resgen.com](http://www.resgen.com))

Seq primer: T3 ET from Amersham  
High quality sequence stop: 467.

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10
11. <b>Feature 11</b>	Source 11
12. <b>Feature 12</b>	Source 12
13. <b>Feature 13</b>	Source 13
14. <b>Feature 14</b>	Source 14
15. <b>Feature 15</b>	Source 15
16. <b>Feature 16</b>	Source 16
17. <b>Feature 17</b>	Source 17
18. <b>Feature 18</b>	Source 18
19. <b>Feature 19</b>	Source 19
20. <b>Feature 20</b>	Source 20
21. <b>Feature 21</b>	Source 21
22. <b>Feature 22</b>	Source 22
23. <b>Feature 23</b>	Source 23
24. <b>Feature 24</b>	Source 24
25. <b>Feature 25</b>	Source 25
26. <b>Feature 26</b>	Source 26
27. <b>Feature 27</b>	Source 27
28. <b>Feature 28</b>	Source 28
29. <b>Feature 29</b>	Source 29
30. <b>Feature 30</b>	Source 30
31. <b>Feature 31</b>	Source 31
32. <b>Feature 32</b>	Source 32
33. <b>Feature 33</b>	Source 33
34. <b>Feature 34</b>	Source 34
35. <b>Feature 35</b>	Source 35
36. <b>Feature 36</b>	Source 36
37. <b>Feature 37</b>	Source 37
38. <b>Feature 38</b>	Source 38
39. <b>Feature 39</b>	Source 39
40. <b>Feature 40</b>	Source 40
41. <b>Feature 41</b>	Source 41
42. <b>Feature 42</b>	Source 42
43. <b>Feature 43</b>	Source 43
44. <b>Feature 44</b>	Source 44
45. <b>Feature 45</b>	Source 45
46. <b>Feature 46</b>	Source 46
47. <b>Feature 47</b>	Source 47
48. <b>Feature 48</b>	Source 48
49. <b>Feature 49</b>	Source 49
50. <b>Feature 50</b>	Source 50
51. <b>Feature 51</b>	Source 51
52. <b>Feature 52</b>	Source 52
53. <b>Feature 53</b>	Source 53
54. <b>Feature 54</b>	Source 54
55. <b>Feature 55</b>	Source 55
56. <b>Feature 56</b>	Source 56
57. <b>Feature 57</b>	Source 57
58. <b>Feature 58</b>	Source 58
59. <b>Feature 59</b>	Source 59
60. <b>Feature 60</b>	Source 60
61. <b>Feature 61</b>	Source 61
62. <b>Feature 62</b>	Source 62
63. <b>Feature 63</b>	Source 63
64. <b>Feature 64</b>	Source 64
65. <b>Feature 65</b>	Source 65
66. <b>Feature 66</b>	Source 66
67. <b>Feature 67</b>	Source 67
68. <b>Feature 68</b>	Source 68
69. <b>Feature 69</b>	Source 69
70. <b>Feature 70</b>	Source 70
71. <b>Feature 71</b>	Source 71
72. <b>Feature 72</b>	Source 72
73. <b>Feature 73</b>	Source 73
74. <b>Feature 74</b>	Source 74
75. <b>Feature 75</b>	Source 75
76. <b>Feature 76</b>	Source 76
77. <b>Feature 77</b>	Source 77
78. <b>Feature 78</b>	Source 78
79. <b>Feature 79</b>	Source 79
80. <b>Feature 80</b>	Source 80
81. <b>Feature 81</b>	Source 81
82. <b>Feature 82</b>	Source 82
83. <b>Feature 83</b>	Source 83
84. <b>Feature 84</b>	Source 84
85. <b>Feature 85</b>	Source 85
86. <b>Feature 86</b>	Source 86
87. <b>Feature 87</b>	Source 87
88. <b>Feature 88</b>	Source 88
89. <b>Feature 89</b>	Source 89
90. <b>Feature 90</b>	Source 90
91. <b>Feature 91</b>	Source 91
92. <b>Feature 92</b>	Source 92
93. <b>Feature 93</b>	Source 93
94. <b>Feature 94</b>	Source 94
95. <b>Feature 95</b>	Source 95
96. <b>Feature 96</b>	Source 96
97. <b>Feature 97</b>	Source 97
98. <b>Feature 98</b>	Source 98
99. <b>Feature 99</b>	Source 99
100. <b>Feature 100</b>	Source 100

```

FEATURES
source
Location/Qualifiers
1..570
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult liv
stage embryos"
/lab_host="XL1-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site
strand cDNA was primed with a Not I - olig
[5']GCACGATGTTCTAGATCGCAGCGCGCCCTTTT
double-stranded cDNA was ligated to Sal I
digested with Not I and cloned into the N
sites of the pSPORT1 vector (BRL). Library
by Matthew Clark (Lehrach lab; ICRF, London
Institut fuer Molekulare Genetik, Berlin).
analysis were selected following oligonucle
hybridization fingerprinting of arrayed cl
zebrafish late somitogenesis (26 ss), adul
embryonic shield stage (5.6 h) libraries.
data were used to computationally cluster
single cDNA from each cluster was chosen f
In some cases multiple members of the same
sequenced to assess clustering parameters
were sequenced additional times to assess
control."
BASE COUNT 150 a 158 c 145 g 117 t
ORIGIN

alignment_scores:
Quality: 770.00 Length: 190
Ratio: 4.208 Gaps: 0
Percent Similarity: 96.316 Percent Identity: 78.421

alignment_block:
US-09-512-581-2 x AI794456 ..

Align seg 1/1 to: AI794456 from: 1 to: 570

712 ProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLysTy 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2 CCAGTACTGCAGCAGCAAGCAAAAGAGGCCCTCTCTCCGACGCCAAGTA 51
728 rAlaLeHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheA 745
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 TGCACATCCAGTGGGTCCAGCCCATGTTTTCACACAGACAGACACACTTG 101
745 laGlntIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGlu 761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 CACAGATATTTGACCTCTGCATTAAGGGTCTAGACACAGACAACATGGAG 151
762 HisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 CAGCTCATCAGCGCTTTTAACGACACTGGGACACTGGCCATCTGCGCCC 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Align seg 1/1 to: AI666974 from: 1 to: 580

```

739 SerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
24 AACAGAGACACACACTTTGCACAGATATTTGAGCCTCTGCATAAGGGTCT 73
755 uasProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyH 772
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
74 AGACACAGACACATCGAGCAGCTCATCATCGCTTTTAAGCACATCGGAC 123
772 isileAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrp 788
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
124 ACTGGCCATGCTGGCCCCAGACAGATTTGCTGCTCCGCTCAAACTTTTA 173
789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGl 805
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 GTGGCCAATTTTCATCGTTAAAGACCTGCTAATGAAGACCGGATTCCTGG 223
805 yLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluT 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 CAAGAAACACCAAGCTGTGGTCCACAGATGACGAAGTCTCTCCTGAAA 273
822 hMetValLysIleGlnAlaIleLysMetValArgTrpLeuLeuGly 838
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 CATCACCAAGATTACGGTCTGAAACTGATGATGCTGGCTGCTCGGT 323
839 MetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuTh 855
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 GTGAACAACAACAGAGTAAATCTGGAACCTCAACTCTCGGATGTTGAC 373
855 rThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerL 872
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 GGCAAATCCGTGAGCAGCGATGGAGATCTGCGGAGCAGGCAAGATGGGA 423
872 ysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLys 888
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 AGCGGACATGCTCTGCTGCGTCTGGCAGCGGTGCGATCTTCGT 473
889 LeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGl 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 CTGCTCAGAGCGGTGCTACCATGATGATCATCACCTGGAGCAGTACCA 523
905 nLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValP 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
524 GCCTCGGCCCTCGTCATCAACGACGAGTGCTACCAGGTGCGACAGGCTN 573
922 heAla 923
|||
574 TCGCT 578

seq_name: gb_est25:AI794456

seq documentation block:
LOCUS AI794456 570 bp mRNA EST
DEFINITION fc44f03.y1 zebrafish WashU MPMG EST Danio rerio cd
sequence.
ACCESSION AI794456
VERSION AI794456.1 GI:5342172
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Cypriniformes; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 570)
Clark.M., Johnson.S.L., Lehrach.H., Lee.R., Li.F.,
S., Hillier.L., Kucaba.T., Martin.J., Beck.C., Wyl
,K., Stepcoe.M., Theising.B., Allen.M., Bowers.Y.,
Swaller.T., Gibbons.M., Pape.D., Harvey.N., Schurk,
Kohn,S., Shin,T., Jackson.Y., Cardenas.M., McCann,R.
and Wilson.R.
WashU zebrafish EST Project 1998

```



/tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker: Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 169 a 101 c 95 g 105 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 786.00 Length: 156  
 Ratio: 5.071 Gaps: 0  
 Percent Similarity: 99.359 Percent Identity: 98.718

alignment\_block:

US-09-512-581-2 x AA282190 ..

Align seg 1/1 to: AA282190 from: 1 to: 470

1075 llelleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPr 1091  
 1 ATCATCATGTCAGAGAGTACTACATACAGTTTGGAAATCTCTAAAGACCC 50

1091 oValLeuProAlaArgPhePheThrGlnProAspLysAsnPhSerAsnT 1108  
 51 GGTACTACAGCTCGTTCTTCACCTACACCTGACAGAAATTCAGTAACA 100

1108 hrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124  
 101 CCAAAATATATCGCTCTCTGAAATGAATCATATTTTCACTCCTCGAANA 150

1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAl 1141  
 151 CCTAAACACCAATGTCTAGGAGCTGTAAACAGCCACTTTCATCAGC 200

1141 agLysGlnSerGlnThrLysSerArgMetGluThrValSerAsnA 1158  
 201 AGGCAAGCAATCTCAGACCAATCATCAGATGGAACCTGTAACCAATG 250

1158 laSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeu 1174  
 251 CAAGCAGCAGCTCAATCCAGCTCTCTGGGAAGTAATAAGGGAGGCTT 300

1175 AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSe 1191  
 301 GATAGTCTGAAATGGATGACAGTGAATGAATGAATGACAAATGCTCTC 350

1191 rProLeuProGlyLysLysSerAspLysArgAspSerAspLeuValA 1208  
 351 ACCTTTGCGGGGAAAAAAGTGACAGAGAGACGACTCTGATCTGTAA 400

1208 rgSerGluLeuLysProArgGlyArgLysLysThrProValThrGlu 1224  
 401 GGTCTGAATGGAGAGCCTAGAGCAGGAGAAAAAACGCTCGTCACAGAA 450

1225 GlnGluGluLysLeuGly 1230  
 451 CAGGAGGAGAAATAGGT 468

seq\_name: gb\_est23:AI666974

seq\_documentation\_block:  
 LOCUS AI666974 580 bp mRNA EST 18-MAY-1999

DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT

FEATURES  
 source

BASE COUNT 145 a 166 c 151 g 117 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 783.00 Length: 185  
 Ratio: 4.399 Gaps: 0  
 Percent Similarity: 96.216 Percent Identity: 79.459

alignment\_block:  
 US-09-512-581-2 x AI666974

fc24a12.y1 Zebrafish Washu MPIMG EST Danio rerio cDNA 5', mRNA  
 sequence.  
 AI666974  
 AI666974.1 GI:4805330  
 EST.  
 Zebrafish.  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Rasbora; Danio.  
 1 (bases 1 to 580)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Bddy  
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood  
 K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,  
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
 and Wilson, R.  
 Washu Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watson.wustl.edu  
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:  
 www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 480.  
 Location/Qualifiers  
 1..580  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone\_lib="Zebrafish Washu MPIMG EST"  
 /sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield  
 stage embryos"  
 /lab\_host="XLI-blue MRF"  
 /note="vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
 strand cDNA was primed with a Not I - oligo(dT)15 primer  
 [5'-PGACTAGTCTAGATCCGAGCGCCGCTTTTCTTTT3'];  
 double-stranded cDNA was ligated to Sal I adaptors (BRL),  
 digested with Not I and cloned into the Not I and Sal I  
 sites of the pSPORT1 vector (BRL). Library was constructed  
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
 analysis were selected following oligonucleotide  
 hybridization fingerprinting of arrayed clones from  
 zebrafish late somitogenesis (26 ss), adult liver or  
 embryonic shield stage (5.6 h) libraries. Fingerprint  
 data were used to computationally cluster cDNAs, and a  
 single cDNA from each cluster was chosen for sequencing.  
 In some cases multiple members of the same cluster were  
 sequenced to assess clustering parameters or single clones  
 were sequenced additional times to assess quality  
 control."







751 Leu.HisLysSerLeuAspProSerAsnLeuGluHisLeuLeuThrProL 767  
 645 CTGTCAGCTAGCTCTGAAGCGCAGAGTCCCAAGACAGCTCATCACTCC 596  
 767 euValThrLeuGlyHisLeuAlaLeuAlaProAspGlnPheAlaAla 783  
 595 TCGTGTCTACTCGCCACATCTTCATGCTGGCCCAAGATCAGTTGCTTCA 546  
 784 ProTrpLysSerTrpValAlaThrPheLeuValLysAspLeuLeuMetAs 800  
 545 CGATGAAGTCCATTGTGGTAACTTCATCGTGAAGGACCTGCTCATGAA 496  
 800 nAspArgLeuProGlyLysLysThrLysLysLeuTrpValProAspGluG 817  
 495 TGACAGATCGGTGGGAAACAAGATGGGAAGCTGTGGACCACTGATG 446  
 817 luValSerProGluThrMetValLysIleGlnAlaIleLysMetVal 833  
 445 AAGTCTACCTGAAGTTCGCTAAGGTGCAGGCCATCAAGTGTGGTG 396  
 834 ArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerTh 850  
 395 CGTGTGTTAGGAATGAAGAACCAACCAATCCAAGTCGGCCCACTCCAC 346  
 850 rLeuArgLeuLeuThrIleLeuHisSerAspLysLeuThrGluG 867  
 345 COTCGCTGCTGTTTCGGCATCTTGGTCAGCGGAGGAGACCTCAGAGC 296  
 867 InGlyLysLysSerLysProAspMetSerArgLeuArgLeuAlaAlaGly 883  
 295 AGAAGAAGATCAGTAATCAGACATGCTTCGGCTGAGGCTGGCCGAGGT 246  
 884 SerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleTh 900  
 245 GGAGCCATTATCAAGTTAGCCAGGAGCTGTGTACCATGACATCATCAC 196  
 900 rLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG 917  
 195 CCTGAACAGTTTTCAGCTTCGGGCTTGTATCAATGACGAGTGTCTACC 146  
 917 InValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933  
 145 AGTTTCGTCAGATCTTGTCTCAGAGCTGCACCTGGCTCTGTCAAGCTG 96  
 934 ArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPr 950  
 95 GTGCTTCCCTCGAGTACTTGGCGGTGTGTGCTGTGTGCTGCAAGGACC 46  
 950 oValLysGluArgArgAlaHisAlaArgGlnCysLeu 962  
 45 GGTGAAGGAGCGCGCGCCCGCCGACAGTGCCTC 9

seq\_name: gb\_est102:BG571128

seq\_documentation\_block:  
 LOCUS BG571128 888 bp mRNA EST 10-APR-2001  
 DEFINITION 602591748F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4714234 5',  
 mRNA sequence.  
 ACCESSION BG571128  
 VERSION BG571128.1 GI:13578781  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 888)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 CONTACT Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLCMI558 row: p column: 11  
 High quality sequence stop: 688.

#### FEATURES

Location/Qualifiers  
 source

1..888  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIH\_MGC\_79"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfII (ggccgctggcc); Site\_2: SfII (ggccattatggcc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 347 a 169 c 223 g 149 t  
 ORIGIN

#### alignment\_scores:

Quality: 806.00 Length: 160  
 Ratio: 5.037 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-512-581-2 x BG571128

Align seg 1/1 to: BG571128 from: 1 to: 888

1232 AspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnAr 1248  
 2 GATGACTTGACTAGTTGGTACAGAACAGAACCTTAAAGGCGAGTCAGCG 51  
 1248 gSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnTrpP 1265  
 52 AAGTCGAAAGAGGCGCATACGCTTCAGATCTGATGAACAGCAGTCGC 101  
 1265 roGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGln 1281  
 102 CTGAGGAAAGAGGCTCAAGAGAGATATATTAGAAATGAAGATGAACAG 151  
 1282 AsnSerProProLysGlyLysArgGlyArgProProLysProLeuGl 1298  
 152 AATAGTCCGCCAAA.AAGGTAAGAGAGCGCCGCCAACCAACCTCTTGG 200  
 1298 yGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlys 1315  
 201 TGGAGGTACACCAAAAGAGAGCCCAACATGAAACTTCTAAAAAGGAA 250  
 1315 erLysLysLysSerGlyProProAlaProGluGluGluGluGlu 1331  
 251 GCAAAAAAATCTGGACCTCCAGCAGGAGGAGGAGGAAGAAGAA 300  
 1332 ArgGlnSerGlyAsnThrGlnLysSerLysSerLysGlnHisArgVa 1348  
 301 AGACAAAGTGGAAATACGACAGCAAGTCCAAAGCAACACACCGAGT 350  
 1348 lSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleG 1365  
 351 GTCAAGGAGGACACAGCAGACAGAGATCTCCTGAATCTAGTCAATTC 400  
 1365 luSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThr 1381



www-bio.lnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

source Location/Qualifiers

1..490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3088558"  
/clone.lib="NCI\_CGAP\_Sub8"  
/lab.host="DH10B (Life Technologies)"  
/note="vector: pMT3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not 1; Site 2: Eco RI; NCI\_CGAP\_Sub8 is a subtracted library derived from NCI\_CGAP\_Sub5. The NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE clone ids 2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI\_CGAP\_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_LIB=NCI\_CGAP\_Lu5  
TAG\_TISSUE=lung  
TAG\_SEQ=CAAC"

BASE COUNT 167 a 96 g 148 t

ORIGIN

## alignment\_scores:

Quality: 816.00 Length: 159  
Ratio: 5.197 Gaps: 0  
Percent Similarity: 98.742 Percent Identity: 98.742

## alignment\_block:

US-09-512-581-2 x BF511856 ..

Align seg 1/1 to: BF511856 from: 1 to: 490

894 CysTyrHisLutlelrhrLeuGluGlnTyrGlnLeuCysAlaLeuAl 910  
|||||  
13 TGTTACCATGAATATCATCATTTAGAACATATATCAGCTATGTCATAGC 62  
910 alleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLeuH 927  
|||||  
63 TATCAACGATGAATGTATCACTAGACAGAGTTTGCCAGAACTTC 112  
927 isLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys 943  
|||||  
113 ACAAGGCGCTTTCCCGTTAGCGCTTCCACTTGAGTATATGCGCAATCTGT 162  
944 AlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgG1 960  
|||||  
163 GCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGAGCTCATGCTAGGCA 212  
960 nCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 977  
|||||  
213 ATGTTTGGTGAATAATATTAATATAGCGGGGAGATATCTGAACGACATG 262  
977 laAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993  
|||||  
263 CAGCTGTAGTGAATAATATTGCTCTCTCTACCAGATATGTTGTTC 312  
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010  
|||||  
313 TATACAATTACCTTTTGGCACATGATGATGATGATGATGATGATGATGATG 362

1010 pileGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluI 1027  
|||||  
363 TATTGAACAACTTAAAGATGCTTAAAGAAATGCTTTGGTTTCTGGA 412  
1027 leLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysMet 1043  
|||||  
413 TATTAAATGGCTAAATAATGAAATAACAGTCACGCTGTTATCAGAAAGATG 462  
1044 valGluAsnIleLysGlnThrLysAsp 1052  
|||||  
463 GTAGAAATATTAAACAAACAAAGAT 489  
seq\_name: gb\_est41:AW013141

## seq\_documentation\_block:

LOCUS AW013141 746 bp mRNA EST 10-SEP-1999  
DEFINITION SRN011SK Winter flounder stomach Pleuronectes americanus cDNA clone sequence.  
ACCESSION AW013141 GI:5861919  
VERSION AW013141.1  
KEYWORDS EST.  
SOURCE Pleuronectes americanus.  
ORGANISM Pleuronectes americanus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes; Pleuronectoidi; Pleuronectidae; Pleuronectes.  
REFERENCE 1 (bases 1 to 746)  
AUTHORS Douglas,S.E., Gallant,J.W., Bullerwell,C.E., Wolff,C., Munholland,J. and Reith,M.E.  
TITLE Winter flounder expressed sequence tags: Establishment of an EST database and identification of novel fish genes  
JOURNAL Marine Biotechnology (1999) In press  
COMMENT Contact: Reith M  
Marine Biology  
NRC Institute for Marine Biosciences  
1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada  
Tel: (902) 426-8276  
Fax: (902) 426-9413  
Email: michael.reith@nrc.ca  
Seq primer: M13 Forward.

## FEATURES

Location/Qualifiers  
1..746  
/organism="Pleuronectes americanus"  
/db\_xref="taxon:8263"  
/clone="SRN011SK"  
/clone.lib="Winter flounder stomach"  
/sex="female"  
/dev\_stage="adult"  
/note="Organ: stomach"  
BASE COUNT 171 a 192 c 209 g 174 t  
ORIGIN

## alignment\_scores:

Quality: 806.00 Length: 247  
Ratio: 3.784 Gaps: 3  
Percent Similarity: 86.235 Percent Identity: 67.611

## alignment\_block:

US-09-512-581-2 x AW013141/rev ..

Align seg 1/1 to reverse of: AW013141 from: 1 to: 746

720 LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI 736  
|||||  
745 CRAAGCCCCAAAGCCAGGGGCTTTCCCAACTGGTAATTCCTCCATGCCCA 696  
736 ePheSerSerLysGlu.....ThrGlnPheAlaGlnIle.PheGluPro 750  
|||||  
695 TTTTAAAAACAACAAAGGAAGTGAAGTCAGCTGGCACAGATTTTGTAGCTT 646



```

|||||
603 TGTGTTATACATAGTCAAGTGTGTGGCAAGGCA..... 640
1085 LeuGluSerProLys... AspProValLeuProAlaAArgPhePhe ThrG 1100
:::|||||::: ||| ||| ||| ||| ||| ||| |||
641 ...GACTCACCAAGGGAACCAAGTCTCCCAANTCAAGTCTTTAACAAC 687
1100 InProAspLysAsnPhe.SerAsnThrLysAsnTyLysLeuProGluMe 1116
:::|||| ||| ::::: ::::: ::::: :::::
688 CGGCTGAGAGAGCCTTCTGTTAGGACAGAGCATATATTATTAGGAACAA 737
1116 tyLysSer PhePheThrProGlyLysProLysThrThrAsnVal 1130
:::|||| ||| ::::: ||| ||| ||| ||| ||| |||
738 CAACAGTGTCTCTGTTGAACGAGAGCAACGCTACGGGGGTA 781

```

seq\_name: gb\_est76:BE551003

```

seq_documentation_block: 535 bp mRNA EST 10-AUG-2000
LOCUS BE551003 7667b03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3233261 3',
DEFINITION mRNA sequence.
ACCESSION BE551003
VERSION BE551003.1 GI:9792695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 475.
Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu25 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
141920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

BASE COUNT 183 a 90 c 108 g 153 t 1 others

```

alignment_scores:
  Quality: 817.00 Length: 170
  Ratio: 4.922 Gaps: 0
  Percent Similarity: 97.647 Percent Identity: 95.294

```

alignment\_block:
US-09-512-581-2 x BE551003

Align seg 1/1 to: BE551003 from: 1 to: 535

```

905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnVa 921
||||| ||||| ||||| ||||| ||||| ||||| |||||
27 CAGCTATGTGCATTAGCTATCAACGATGATGCTNATCAAGTAGACAAGT 76
921 lPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG 938
||||| ||||| ||||| ||||| ||||| ||||| |||||
77 GTTGGCCCAAGAAACTTAC.AAAGGCGCTTTCCCGTTTACGGCTTCCATT 125
938 luTyrrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
||||| ||||| ||||| ||||| ||||| ||||| |||||
126 AGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTTAAGGAGA 175
955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgG 971
||||| ||||| ||||| ||||| ||||| ||||| |||||
176 AGAGCTCATGCTAGGCAATGTTGGTGAATAATATAAATGTAAGCGGGA 225
971 uTyrrLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuP 988
||||| ||||| ||||| ||||| ||||| ||||| |||||
226 GTATCTGAAGCAGCATGCGAGCTGTAGTGAATAATATTGTCTCTCTAC 275
988 roGluTyrrValValProTyrrThrIleHisLeuLeuAlaHisaspPro 1004
||||| ||||| ||||| ||||| ||||| ||||| |||||
276 CAGAGTATGTTGTCATATACATACATTCACCTTTGGCAGCATGACCCAG 325
1005 TyrrValLysValGlnAspIleGlnLeuLysAspValLysGluCysLe 1021
||||| ||||| ||||| ||||| ||||| ||||| |||||
326 TATGTCAAGTACAGGATATTGAACAACCTTAAGATGTTAAGAATGCT 375
1021 uTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHis 1038
||||| ||||| ||||| ||||| ||||| ||||| |||||
376 TTGGTTTGTCTGGAATATTATGCGTGAATAATGAAATAACAGTCAGC 425
1038 lPheIleileArgLysMetValGluAsnIleLysGlnThrLysaspAla 1054
||||| ||||| ||||| ||||| ||||| ||||| |||||
426 CTTTATATCAGAAAGATGGTCGAAAATTTTAAACAAACAAAGATGCC 475
1055 GlyProAspAspAlaLysMetAsnGluLysLeuTyrrThrValCysasp 1071
||||| ||||| ||||| ||||| ||||| ||||| |||||
476 GCACCATGATGATGCAAAATGATGAAAACTGTACACTGTGTGTGATCT 525
1071 lAlaMetAsn 1074
||||| |||||
526 TGCCATGAAT 535

```

seq\_name: gb\_est88:BF511856

```

seq_documentation_block:
LOCUS BF511856 490 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-B14-aps-d-12-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088558 3', mRNA sequence.
ACCESSION BF511856
VERSION BF511856.1 GI:11595154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:

```



```

1154 ThrValSerAsnAlaSerSerSerAsnProSerProGlyArgII 1170
|||||
480 ACTGTAGCAATGCAAGCACACACTCAATCCAAAGCTCTCTCTGGAAGAT 431
|||||
1170 eLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspT 1187
|||||
430 AAAGGGGAGGCTTGATGTTCTGAAATGGATCACAGTGAAATGAAGATT 381
|||||
1187 yThrMetSerProLeuProGlyLysSerSerAspLysArgAspAsp 1203
|||||
380 ACACAATGCTTCCACCTTCCGCGGAAAAAAGTGCACAGAGAGACGAC 331
|||||
1204 SerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysTh 1220
|||||
330 TCTGATCTTGAAGTCTGAATTTGGAGAACCCCTAGAGCGAGGAAAAAAC 281
|||||
1220 rProValThrGluGlnGluLysLeuGlyMetAspLysLeuThrLysL 1237
|||||
280 GCCCGTCACAGACAGGAGGAGAAATTAGGTATGGATGACTTGCATAAGT 231
|||||
1237 euValGlnGluLysProLysGlySerGlnArgSerArgLysArgGly 1253
|||||
230 TGTGTACAGGAACAGAACCTTAAGGCGAGTCAGCGAAGTCGGAAGAGGC 181
|||||
1254 HisThrAlaSerGluSerAspGluGlnGlnTrpProGluLysArgLe 1270
|||||
180 CATACGGCTTCAGAACTGTATGAAACAGCAGTCGGCTGAGGAAAAAGGCT 131
|||||
1270 uLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysL 1287
|||||
130 CAAAGAAGATATATTAGAAATCAAGATGAACAGATAGTCGCCAAAAA 81
|||||
1287 ySGLysArgGlyArgProProLysProLeuGlyGlyThrProLys 1303
|||||
80 AGGGTAAAGAGCGCCAGCACCAACCAACCTCTTGGTGGAGGTACACCAAAA 31
|||||
1304 GluGluProThrMetLysThrSerLysLys 1313
|||||
30 GAAGAGCAAAAATGAAGAACTTCTAAAAAA 1

```

seq\_name: gb\_est97:BG176414

```

seq_documentation_block:
LOCUS BG176414 910 bp mRNA EST 06-FEB-2001
DEFINITION 602338082F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4460901 5',
mRNA sequence.
ACCESSION BG176414
VERSION BG176414.1 GI:12683026
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10263 row: d column: 22
High quality sequence stop: 628.
Location/Qualifiers
1..910
/organism="Mus musculus"
/strain="FVB/N"

```

# FEATURES

source

```

/db_xref="taxon:10090"
/clone="IMAGE:4460901"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 276 a 199 c 225 g 210 t
ORIGIN

```

alignment\_scores:  
Quality: 827.50 Length: 268  
Ratio: 3.662 Gaps: 8  
Percent Similarity: 84.328 Percent Identity: 63.433

alignment\_block:  
US-09-512-581-2 x BG176414 ..

Align seg 1/1 to: BG176414 from: 1 to: 910

```

874 AspMetSerArgLeu ArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
|||||
3 GATATGTCCTCGCTGGGATTAGCTGCTGTAGTGCCTAATGAGGCTTG 52
|||||
890 laGlnGluProCysTyrHisGluIleThrLeuGluGlnTyrGlnLeu 906
|||||
53 CTCAGGAACCTTGCTACCATGAAATATTACCACCAGAACAGTTTCAGCTC 102
|||||
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
|||||
103 TGTGCACCTGGTTATTATGATGAGTGTCTACCAAGTAAGGCAGATATTGC 152
|||||
923 aGlnLysLeuHisLysGlyLeuSer ArgLeuArgLeuProLeuGluTyr 939
|||||
153 CCAGAAAGCTTCATAAGGCACCTTGGTGAAGTTGCTCTCCACCTGGAGTAT 202
|||||
940 MetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAl 956
|||||
203 ATGCCATCTTTGCTTGTGTGCAAGACCCCTGTGAAGGAAAGGAGAGC 252
|||||
956 aHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrL 973
|||||
253 ACACGCTCGGCAGTGTGTGTTAAAGAACATCAGCATCCGCGAGGAGTACA 302
|||||
973 euLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuProGlu 989
|||||
303 TCAACAGAACCCCATGCGCCACCTGAGAAATATTGTCACTGCTGCTGAA 352
|||||
990 TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
|||||
353 TATGTGGTTCATACATGATTACCTCTCTAGCCCATCATCTGATTTTCAC 402
|||||
1006 LysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1023
|||||
403 ACGTCCACAAGATGTTGATCAACTTCGTATATAAAGAGTGCCTGTGGT 452
|||||
1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPhe 1039
|||||
453 TTTATGTTGAAGTCTTAATGACAAAGAAATGAAACACACCCATGCATTC 502
|||||
1040 IleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPr 1056
|||||
503 ATGAAAAAGATGCGAGAGATATCAAGCTAACACAGACACGCCAGTCTCC 552
|||||
1056 oAspAspAlaLysMetAsnGluLysLeuTyr.....ThrV 1068
|||||
553 CGATGAATCCAGACAAATGAAAACTTATACGGTTTCTGACGTGGCTCTG 602
|||||
1068 alCysaspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084

```



Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 58 row: L column: 16

Seq primer: ATTATGCTGACACTATAG.

Location/Qualifiers

1..560

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

BASE COUNT	193 a	96 c	121 g	150 t
------------	-------	------	-------	-------

ORIGIN

alignment\_scores:

Quality:	830.00	Length:	186
Ratio:	4.637	Gaps:	0
Percent Similarity:	96.237	Percent Identity:	82.258

alignment\_block:

US-09-512-581-2 x BE756042 ..

Align seg 1/1 to: BE756042 from: 1 to: 560

```

334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
|||||
1 GTGAAATTTGCCAGTCACCTGTTTAAATGAATCAATCCAGATTTAGCAAAGA 50
|||||
350 pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367
|||||
51 TCTCACAGATATTGAAAGTTTCGATCCCATGATCCAGGAAGACTATTTC 100
|||||
367 rgHisAspValIleValSerIleValThrAlaAlaLysAspIleLeu 383
|||||
101 GTCATGATCTTATTGTTACTATAATAAAGTCTGCCAAAAGACCTTGGC 150
|||||
384 LeuValAsnAspHisLeuLeuAsnPhelArgGluArgThrLeuAspIly 400
|||||
151 TTAGTGAATGATCAGCTACTTGGCTTTGTAAGAGAAGAACACTGGACAA 200
|||||
400 sArgTTPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrL 417
|||||
201 ACGGTGGCAGATGAAGAAAGAGCTATGATGGTCTGGCTCAGCTTTATA 250
|||||
417 ySLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaLysGlnIle 433
|||||
251 AGAAATACTGTCTTCATGCTGAAGCAGGAAGAAAGCTGCAGAGAAAGTC 300
|||||
434 AlaTTrpIleLysAspIleLeuHisIleTyrTyrGlnAsnSerIleAs 450
|||||
301 AGCTGGATAAAGGACAACTTTTGCATATATATATCAAAATAGCATCGA 350
|||||
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
|||||
351 TGCAAAACTATTGTTAGAAATACTTCTCAATATCTTGTCCTCCACCA 400
|||||
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
|||||
401 ACCTGGAACACAGAAGAGAGAATGAATGCTTGTATTATTATATGCTACT 450
|||||
484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500

```

```

|||||
451 TTGGATCCAAATGCTGTCAAAGCTCTCAATGAATGTGGAAGTGCAGAA 500
|||||

```

```

500 nLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProL 517
|||||

```

```

501 CATGCTTAGAAGTCACTGTACGAGAACTGTTGGATTGTCACACAGCAGCTA 550
|||||

```

```

517 ySThrAsp 519
|||||

```

```

551 CATCAGAA 558
|||||

```

seq\_name: gb\_est23:AI655429

seq\_documentation\_block:

LOCUS	AI655429	480 bp	mRNA	EST	14-DEC-1999
DEFINITION	ts98f06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2239331 3',				
	mRNA sequence.				

ACCESSION AI655429 GI:4739408

VERSION AI655429.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 480)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

Clone Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1048 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 455.

FEATURES

source

1..480

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2239331"

/clone\_lib="NCI\_CGAP\_GC6"

/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Plasmid DNA from the normalized library

NCI-CGAP\_GC4 was prepared, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from a pool of 5,000 clones made

from the same library (cloneids 1257096-1258631,

1469064-1470983, and 1475592-1476743). Subtraction by

Bento Soares and M. Fatima Bonaldo."

BASE COUNT 82 a 123 c 84 g 191 t

ORIGIN

alignment\_scores:

Quality:	828.00	Length:	160
Ratio:	5.208	Gaps:	0
Percent Similarity:	99.375	Percent Identity:	99.375

alignment\_block:

US-09-512-581-2 x AI655429/rev ..

Align seg 1/1 to reverse of: AI655429 from: 1 to: 480



```

seq_name: gb_est102:BG574948

seq_documentation_block:
LOCUS BG574948 886 bp mRNA EST 10-APR-2001
DEFINITION 602598035F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4706697 5',
mRNA sequence.
ACCESSION BG574948
VERSION BG574948.1 GI:13582601
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 886)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10575 row: f column: 10
High quality sequence stop: 637.
FEATURES
location/Qualifiers
1..886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4706697"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: breast; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 282 a 167 c 221 g 216 t
ORIGIN

alignment_scores:
Quality: 850.00 Length: 279
Ratio: 3.542 Gaps: 6
Percent Similarity: 86.022 Percent Identity: 65.233

alignment_block:
US-09-512-581-2 x BG574948 ..
Align seg 1/1 to: BG574948 from: 1 to: 886
400 LysArgTIPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyr 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 AAACGGTGGCGATAGAAAAGAGCTATGATGGTCTGGCTCAGCTTTA 78
416 rLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnI 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 TAAGAAATACCTGCTTCATGGTGAAGCAGGAGGAAGCTGCAGAGAAAG 128
433 leaLatrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 TCAGCTGGATAAGGACAACTTCGCATATTTATTATCAGACAGCATTT 178
450 AspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHi 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 GAGCACAACCTGTTGGTAGAGAAACTTTTGCCTCAGTATCTTGTCCCCA 228
466 sAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlat 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq\_name: gb\_est78:BE756042

seq\_documentation\_block:

LOCUS BE756042 560 bp mRNA EST 15-SEP-2000  
DEFINITION 210069 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BE756042

VERSION BE756042.1 GI:10170034

KEYWORDS EST.

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 560)

AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid

,W.W. and Keeler,J.W.

TITLE Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

JOURNAL Unpublished (2000)

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA



[illegible]

```

seq_name: gb_est95:BG016343

seq_documentation_block:
LOCUS      BG016343      662 bp      mRNA      EST      24-JAN-2001
DEFINITION de59b11.yl Kirschner embryo St10 14 Xenopus laevis cDNA clone
IMAGE:3515876 5' similar to TR:Q9Y4D4 Q9Y4D4 KIA00648 PROTEIN ;,
mRNA sequence.
ACCESSION  BG016343
VERSION    BG016343.1 GI:12469488
KEYWORDS  EST.
SOURCE    African clawed frog.
ORGANISM  Xenopus laevis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
           Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 662)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
           , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
           ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
           Waterston,R. and Wilson,R.
TITLE     Washu Xenopus EST project, 1999
JOURNAL   Unpublished (1999)
COMMENT   Other_ESRs: de59b11.xl
           Contact: Sandy Clifton, Ph.D.
           Washu Xenopus EST project, 1999
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Library constructed by M. Kirschner (Harvard Medical School). DNA
           Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: Xenopus clones from this library are available
           through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
           Seq primer: -408P from Gibco
           High quality sequence stop: 478.
           Location/Qualifiers
             1..662
               /organism="Xenopus laevis"
               /db_xref="taxon:8355"
               /clone="IMAGE:3515876"
               /clone_lib="Kirschner embryo St10 14"
               /tissue_type="pooled embryos (stage 10-14)"
               /lab_host="DHI0B (phage-resistant)"
               /notes="Vector: pCS2+; Site_1: NOTI; Site_2: SalI;

FEATURES
             source

```

```

BASE COUNT      185 a      144 c      156 g      177 t
ORIGIN

alignment_scores:
    Quality:      852.00      Length:      220
    Ratio:        4.239      Gaps:        0
    Percent Similarity: 91.364      Percent Identity: 75.909

alignment_block:
US-09-512-581-2 x BG016343  ..

Align seg 1/1 to: BG016343 from: 1 to: 662

745 AlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuG1 761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3 GCTCAAAATTTTTCAGGCCCCTCAGCAGAAAGCTTAACACGCAGATGTACCACA 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

761 uHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlap 778
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 ACAATTAGTATTACACCTTTAGTCTCTTTGGGFCACATTTTCGATGCTGGCTC 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

778 rOAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleVal 794
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 CTGACCAATTTGCTTCTCCAATGAATCTGTTGTGGCAAACTTCATTGTGA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

795 LysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLys 811
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 AAGGATCTGCTGATGAACGATCGGTCTAATGGTGCACAAAGATGGCAAACT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

811 uTrpValProAspGluGluValSerProGluThrMetValLysIleGlnA 828
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 GTGGTCCCAGATGAGGAAGTGTCTCCAGAAGTCTTCTTAAGGCCCAAG 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

828 laIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSer 844
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 CAATAAAGCTATTAGTTCGTTGTTGTTGCTGGGTATGAAAAACAATCAGTCA 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

845 LysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAs 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 AAGTCGGCTAAACTCCACTCTACGGTTATTATCAGCTATGCTTGTGAGCGA 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

861 pGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgL 878
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353 GGGAGATCTGACAGACAGAAAAAATAATAGTAATCCGACATGTCCCGGT 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

878 euArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCys 894
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 TAAGGTGTGCTGTGCGCTGCTATCATGAAGCTTGCCAGAAACCGTGC 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

895 TyrHisGluIleIleThrLeuGluGlnTyrglnLeuCysAlaLeuAlaI 911
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453 TATCATGAATCATATTACTCCAGACAGCTTCAGCTCTGTGCTTGGTTAT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

911 eAsnAspGluCysTyrglnValArgGlnValPheAlaGlnLysLeuHisL 928
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503 CAATGATGAATGCTACCAGGTGCGGCAGATCTTTGCACAGAAGCTTCATA 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

928 ysGlyLeuSerArgLeuArgLeuProLeuGluTyrrMetAlaIleCysAla 944
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
553 AGGCTCTAGTTAAACTGCAGCTCCCTCTAGAGTATATGGCTATATTTCGA 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

945 LeuCysAlaLysAspProValLysGluArgArgAlaHisAlaAraGlnC 961
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
603 CTGTGCGCCATAGATCCAGTAAGAAGGCGGCGCATGCCAGACAGTGC 652
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

961 sLeuValLys 964
|||||:|||||
653 CTTACTCAAG 662

```



```

|||||
2  TTTTATTACTTACGAAACATTCGTTGGGCAAGGCATATAGCATATGCTT 51
145 eGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuP 162
52  TGAGTTAGAGATAGCAATGAATTTTACCCAGCTATACAGAACCTTAT 101
162 heSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetVal 178
102 TTTTCAGTTATAACATGGCCACAATCAGAAAGTCCATATGCACATGGTA 151
179 AspLeuMetSerSerIleLeuGluGluGluGluGluGluGluGluGlu 195
152 GACCTTATGAGCTCTATTATTCTGAAGGTGATACAGTGTCTCAGGAGCT 201
195 uLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnL 212
202 TTTTGGATACGGATTAGTAAATCTGTACCTGCTCATAAGAATTTAAGCA 251
212 yGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAla 228
252 AGCAGCATATGATTGGCAAGGCTTTACTGAAGGACAGCTCAAGCT 301
229 IleGluProTyrIleThrPhePheAsnGlnValLeuMetLeuGlyLys 245
302 ATTGAGCCATATATTACCAATTTTAAATCAGGATCTGATGCTTGGAA 351
245 sThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluL 262
352 AACATCTATCAGCGATTGTGTCAGAGCATGTCTTTGACTTAATTTTGAGC 401
262 euTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
402 TCTACAAATATTGATGTCATTTCTGCTCTCTGCTTTACCCAGCTTGAA 451
279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLe 295
452 TTTTAACTAAAGAGCAATGATATAGGAGCGCTACAAGCTCATAACT 501
295 uLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnL 312
502 ACTGGC....ATGTTAGTGGCAAGGATTCAGAAAGTGGCTTCT..... 540
312 ysProLeuTyrGlnCysTyrLeuGlyArgPheAsnAspIleHisValPro 328
541 .....CCA 543
329 IleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPr 345
544 ATCCGCTGGAATGTGTG...TTTCGCTAGCATTCCTCATGAACCATCC 590
345 OAspLeuAlaLys 349
591 TGATTTAGCACAA 603

```

seq\_name: gb\_est100:BG391029

```

seq_documentation_block: 885 bp mRNA EST 12-MAR-2001
LOCUS BG391029
DEFINITION 60241772F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537108 5',
mRNA sequence.
ACCESSION BG391029
VERSION BG391029.1 GI:13284477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM10461 row: 1 column: 05
High quality sequence stop: 695.
FEATURES
    source
        1..885
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4537108"
                /clone_lib="NIH_MGC_92"
                /tissue_type="embryonal carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 277 a 167 c 202 g 239 t
ORIGIN

```

```

alignment_scores:
    Quality: 884.00      Length: 259
    Ratio: 3.794        Gaps: 3
    Percent Similarity: 89.961      Percent Identity: 72.973
alignment_block:
US-09-512-581-2 x BG391029 ..
Align seg 1/1 to: BG391029 from: 1 to: 885
761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAl 777
|||||
6  GAACAACCTTATAACTCCATGATGTTTC.ATGGCCACATTTCTATGTAGC 54
777 aProAspGlnPheAlaAlaProTyrPlySerTrpValAlaThrPheIleV 794
|||||
55 ACCAGATCAGTTTGTCTCCCAATGAATCTGTAGTAGCAAAATTTATTG 104
794 aLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLys 810
|||||
105 TGAAGATCTGCTTAATGAATGACAGGTCAACAGGTGAAAGAAATGCAAAA 154
811 LeuTrpValProAspGluGluValSerProGluThrMetValLysIleG 827
|||||
155 CTGTGCTCTCCAGATGAAGAGTTTCCCTGGAAGTACTAGCAAGGTACA 204
827 nAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 844
|||||
205 GGCAATTAACCTTCTGTAGAGTGGCTGTG. GGTATGAAAAACAACAGT 253
844 erLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
|||||
254 CTAAATCTGCAATTAACCCCTTCGGTTATTATCAGCGATGTGGTTAGT 303
861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerAr 877
|||||
304 GAGGGTGACCTGACAGAGAAAGAGATCATGAATCTGATATGCTCG 353
877 gLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluPro 894
|||||
354 CTTGCGATTAGCTGCTGGTAGTCCCAATGAAGCTTGCTCAGGAACCTT 403
894 yTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAla 910
|||||
404 GTTACCATGAATATTATACCCCAAGACAGTTTCAGCTCTGTGCACTGTT 453

```







```

:::|||||
602 ACCAGGTTCTCGGCGATGATGAGAACTTCGGTCTCAGTTGGAGTTATTA 651
567 ValSerProThrCysSerCysLysGlnAlaGluClyCysValArgGlu 582
552 ATTAGCCCAACCTGTACTTGCACAAACAGCAGATATCGGTGTGAGAGAA 699

```

seq\_name: gb\_est92:BF789050

```

seq_documentation_block:
LOCUS BF789050 902 bp mRNA EST 12-JAN-2001
DEFINITION 602104907F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223040
5', mRNA sequence.
ACCESSION BF789050
VERSION BF789050.1 GI:12094086
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9810 row: n column: 01
High quality sequence stop: 655.
Location/Qualifiers
1..902
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

```

```

BASE COUNT 274 a 156 c 233 g 239 t
ORIGIN
1..902
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

```

```

alignment_scores:
Quality: 899.00 Length: 262
Ratio: 3.777 Gaps: 5
Percent Similarity: 90.840 Percent Identity: 72.901

```

alignment\_block:

US-09-512-581-2 x BF789050

Align seg 1/1 to: BF789050 from: 1 to: 902

```

208 LysAsnLeuAsnLysGlnAlaTyRAspLeuAlaLysAlaLeuLysAr 224
|||||
2 AAGANCTTAATAAAGTCTTCGACCTTGCACAAAGTCTACTGAAAG 51
224 gThAlaGlnAlaLleGluPro...TyRleThrThrPhePheAsnGlnVa 240
|||||
52 GACAGTCCAGACATTGAAGCATGTAT...GCCAATTTTTCATCAAGT 98
240 lLeuMetLeuGlyLysThrSerLeSerAspLeuSerGluHisValPheA 257
|||||
99 CCTGGTCTGGCAGATCATCAGTCAGCAGCCTGTCTGAACACAGTATTG 148

```

```

257 spLeuileLeuGluLeuTyRAsnIleAspSerHisLeuLeuLeuSerVal 273
|||||
149 ATCTGATTCAGGAACCTTTTCTCT.ATCGATCCTCAGTTACTGTTATCTGTC 197
274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLe 290
:::|||||
198 ATGCCACAGCTTGAATTCAAACTGAAGCAACGATGGTGAAGAACGCT 247
290 uGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluL 307
|||||
248 AGCTGTGGTTCGACTCTCGCAAAATTTTTCGGCTCTAAAGATTTCAGATT 297
307 euAlaSerGlnAsnLysProLeuTyrGlnCysTyRLeuGlyArgPheAsn 323
|||||
298 TAGCACACAGATCGGCTCTCTGGCAGTCTTCTTGGCGGATTAAAT 347
324 AspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCy 340
|||||
348 GACATTCATCTCTCTGTGAGTTAGAAAGTGTGAAGTTTGCAGCCACTG 397
340 sLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyRLeuLysV 357
|||||
398 TTTGATGTAATCAACCCAGACTTAGCGAAGGATCTGACAGAAATATTGAAG 447
357 alArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
|||||
448 TTAGGTTCATGATTCACAGAAAGCCATTCTCATGATGTTATTGTTACT 497
374 Ile.ValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuL 390
|||||
498 ATAAATAACAGCTGCCAAAAGAGACCTTGCTTAGTAATCATCAGTTGC 547
390 euAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLys 406
|||||
548 TTGGCTTTTGTGAGGAAAGGACACTGGAATAACGGTGGCGAGTAAGAAA 597
407 GluAlaMetMetGlyLeuAlaGlnIleTyRLeuLysLysTyRAlaLeuGlnSe 423
|||||
598 GAAGCCATGATGGTCTGGTCTGATGTC.TATAAGAAATACACTGCTTCATGG 646
423 rAlaAlaGlyLysAspAla...AlaLysGlnIleAlaTrpIleLysAspL 439
:::|||||
647 GGAAGCAGCAGAGAGGAAAGCCGGGGAGAAAGTCAAGTGGATTAAAGGACA 696
439 yLeuLeuHisIleTyRLeuGln.AsnSerIleAspAspArgLeuLeuVa 455
|||||
697 AGTGGTGGATATCTACTATCAAGACAGCTTTGTTGACAAATATGTTG 746
455 lGluArgIlePheAlaGlnTyRMetValPro 465
|||||
747 TGAGGAAATTTTGGTCAGATTTTGGTCCCA 777

```

seq\_name: gb\_est28:AL045878

```

seq_documentation_block:
LOCUS AL045878 738 bp mRNA EST 29-FEB-2000
DEFINITION DKFZp434J056_r1 434 (synonym: htes3) Homo sapiens cDNA clone
DKFZp434J056 5', mRNA sequence.
ACCESSION AL045878
VERSION AL045878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 738)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Koehrer, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Koehrer K
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert

```



```

943 CysAlaLeuCyAlaLysAspProValLysGluArgAlaHisAlaAr 959
|||||
927 TTTGCTTTTGTGCGCAAGACCTGTGAAAGAAAGAGAGACACAGCTCG 976
|||||
959 glnCysLeuValLysAsnIleAsnValArgGluTyrLeuLysGlnH 976
|||||
977 GCAGTGTGTTGTTAAAGAACATCAGCATCCGACGGAGTACATCAACACA 1026
|||||
976 isAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValVal 992
|||||
1027 ACCCCATGCCACTGAGAATATTGTCACCTGCTGCTGAATATGCGTT 1076
|||||
993 ProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValG1 1009
|||||
1077 CCATACATGATTCACCTCCTAGCCCATGATCTGATTTCCACACGGTCACA 1126
|||||
1009 naspIleGluInLeuLysAspValLysGluCysLeuTrpPheValLeuG 1026
|||||
1127 AGATGTTGATCAACTCGGTGATATAAAGAGTGCCTGTGTTATGCTTG 1176
|||||
1026 luIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLys 1042
|||||
1177 AAGTCTTAATGACAAAGATGAACAACACAGCCATGCATTCATGAAGAAG 1226
|||||
1043 MetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAl 1059
|||||
1227 ATGGCAGAGAATATCAAGCTAACACAGAGCCGACCTCCCGATCAATC 1276
|||||
1059 aLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1076
|||||
1277 CAAGACAAATGAAGAACTTTATACGGTTTGTACGGTGGCTCTGTGTGTTA 1326
|||||
1076 leMetSerLysSerThrTyrSerLeuGluSerProLysAspProVal 1092
|||||
1327 TAAATAGTAAAGTCTGTTGTCANTGCAGACTCACCAAGGACCCAGTC 1376
|||||
1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThr 1108
|||||
1377 CTCCCAATGAAGTTCTTTACACAGCCTGAAAGGTAATTTCCCTCACA 1424
|||||
seq_name: gb_est101:BG473117

```

```

seq_documentation_block: 701 bp mRNA 21-MAR-2001 EST
LOCUS BG473117
DEFINITION 602515089F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:464687 5',
mRNA sequence.
ACCESSION BG473117
VERSION BG473117.1 GI:13405392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI422 row: a column: 24
High quality sequence stop: 683.
Location/Qualifiers
1. 701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:464687"
/clone_lib="NIH_MGC_16"

```

```

/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 233 a 128 c 166 g 174 t
ORIGIN

```

## alignment\_scores:

```

Quality: 919.00 Length: 234
Ratio: 4.177 Gaps: 2
Percent Similarity: 94.017 Percent Identity: 75.641

```

## alignment\_block:

```
US-09-512-581-2 x BG473117 ..
```

```
Align seg 1/1 to: BG473117 from: 1 to: 701
```

```

351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
|||||
2 CTCACAGAATATTAAAGCTTAGATCAGCTGATCCAGAGAAGCTATTTCG 51
|||||
367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
|||||
52 TCATGATGTCATTGTTTACTATATAACAGCTGCCAAGAGGGACCTGGCCT 101
|||||
384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
|||||
102 TAGTAATGATCAGCTGCTGGCTTTGTAAGGAAAGAACACACTGGATAA 151
|||||
401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
|||||
152 CGGTGGCAGAGTAAGAAAGAAAGCTATGATGGGTCTGGCTCAGCTTATAA 201
|||||
417 sLYSTYrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGln Ile 433
|||||
202 GAAATACTCTCTCATGTTGTAAGCAGGAAAGGAAGCTGCAGACGAAAGTC 251
|||||
434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
|||||
252 AGCTGGATAAAGSACAAACTTCTGCATATTATTATCAGAACACGATTGA 301
|||||
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 457
|||||
302 CGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCGCCGAGA 351
|||||
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
|||||
352 ACCTGGAAACAGAGAGAGAAATGCTTATTATTACTTATATGCTAGT 401
|||||
484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500
|||||
402 TTGGATCCAAATGCTGTAAGAGCTCTCAACGAAATGTGGAAGTGCAGAA 451
|||||
500 nLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProL 517
|||||
452 CATGCTTCGGAGCAGTATACGGAACTATTGGATTTCACACAGCAGCTTA 501
|||||
517 yThrAspAlaSerValLysAlaIlePhe SerLysValMetValIlePh 533
|||||
502 CATCAGAGGCTACTCTTCTGCCATGTTCCGAAAAAAGCTGATGACCATAGC 551
|||||
533 rArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 550
|||||
552 AAAGAATCTGCCTGTACCCCGGGAAGACACAAAGATTACGTGAAGAATTTA 601
|||||
550 hrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
|||||

```











```

964 sAsnIleAspValArgGluTyrLeuLysGlnHisAlaAlaValSerG 981
|||||
206 AAATATAATGTAAGCGGAGTATCTGACGACGATGCGACCTGTAGTG 255
|||||
981 lLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
|||||
256 AAAAATTATTGTCTCTCTACGAGATGTTGTTCCATACAAATTCAC 305
|||||
998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLe 1014
|||||
306 CTTTGGCACATGACCCAGATATGTCAAAGTACAGGATATTGAACAAC 355
|||||
1014 uLysAspValLysGluCysLeuTyrPheValLeuGluLeuMetAlaL 1031
|||||
356 TAAAGATGTTAAAGAAATGCTTTGGTTGTCGGAATATTATAGGCTA 405
|||||
1031 yAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
|||||
406 AAAATGAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAATATT 455
|||||
1048 LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluL 1064
|||||
456 AAACAACAAGATGCCCAATGACCATGATGATGACAGAAATGAATGAAA 505
|||||
1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1081
|||||
506 ACITGACATGTTGTGATGTCGATGTCATGATATCATCATGTCACAGAG 555
|||||
1081 hrThrTyr.SerLeuGluSer 1087
|||||
556 CTACATACNAGTTTGAATCT 576
|||||

```

seq\_name: gb\_est101:BG532020

```

seq_documentation_block:
LOCUS BG532020 735 bp mRNA EST 03-APR-2001
DEFINITION 602561049F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4698900 5',
mRNA sequence.
ACCESSION BG532020
VERSION BG532020.1 GI:13523558
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1530 row: a column: 13
High quality sequence stop: 727.
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4698900"
/clone_lib="NIH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcccttcggcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
and 3' adaptors were used in cloning as follows: 5'

```

adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."

BASE COUNT 223 a 141 c 135 g 236 t  
ORIGIN

#### alignment\_scores:

Quality: 944.00 Length: 246  
Ratio: 4.196 Gaps: 3  
Percent Similarity: 91.463 Percent Identity: 78.049

#### alignment\_block:

US-09-512-581-2 x BG532020 ..

Align seg 1/1 to: BG532020 from: 1 to: 735

```

34 ArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerG1 50
|||||
5 AGACTACTGATGTAGTGAACCTTTATGGATATGGATCAGGACTCAGA 54
|||||
50 uGluGluLys.GluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
|||||
55 AGATGAAAGACAGACAGTATCTCCACATAGCCTTGCATCTTGCATCTGAA 104
|||||
67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCy 83
|||||
105 TTCTCTCAGGAACCCCAATAAGATGTGCTCTCCTTGTAGCATGTTG 154
|||||
83 sLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerP 100
|||||
155 TTTGGCTGATATCTTCGTATCTATGCCCCAGAGCTCCATATACTTCCC 204
|||||
100 roAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
|||||
205 ATGATAAACCTTAAGGACATATTTTGTGTTATTACCAGACAATTAAGAAGT 254
|||||
117 LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuG1 133
|||||
255 TTGAGAGTACAAAGAGTCCACAGTTTAATAGATACTTTTATTATTATAGA 304
|||||
133 uAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAsps 150
|||||
305 GAATTTAGCTTGGTTAAATCATATAACATCTGCTTTGAATTTGGAAGATT 354
|||||
150 erAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166
|||||
355 GCAATGACATTTTATTACGCTTTTATAGAACTCTCTTCTCAGTGATCAAC 404
|||||
167 AspGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSe 183
|||||
405 AATAGCCACANTAGAGAGGTACAAATGCATGCTAGATTGTGATGATTC 454
|||||
183 rIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 200
|||||
455 TATCATCATGGAAGTGTGAGTACTTACTCAAGAAATATTGGACTCCATTC 504
|||||
200 euValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAsp 216
|||||
505 TTATTAACCTTCATCTCTGACACATAAGAACTTAATAAACAGTCTCTTTGGC 554
|||||
217 LeuAlaLysAlaLeuLeu.LysArgThrAlaGlnAlaIleGluProTyrI 233
|||||
555 CTTGCAAAAGTGTATTGACACAGACAGTCCACAGTATTATGAGCATGCA 604
|||||
233 lerThrThrPheAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
|||||
605 TTGCCAATTTTTCATCAATCAAGTCTCTGTGCTGGGAAGATCATCAGTACGT 654
|||||

```



```
seq_documentation_block:
LOCUS      AW770571          578 bp    mRNA
DEFINITION hl86f11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3008877 3'
```



normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M\_S4, NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1, NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996).

BASE COUNT 193 a 113 c 128 g 174 t  
ORIGIN

alignment\_scores:  
Quality: 969.00 Length: 186  
Ratio: 5.238 Gaps: 0  
Percent Similarity: 99.462 Percent Identity: 99.462

alignment\_block:  
US-09-512-581-2 x BF471938 ..

Align seg 1/1 to: BF471938 from: 1 to: 608

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17  
|||||  
50 ATGGCTCATTCACAGACAGGACACGATGGGAAATTTACTTACCTCC 99  
|||||  
17 oGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArg 34  
|||||  
100 TGGAGTCAAGGAATCTCAGATAAAATCTCTAAAGAGGAGATGCTGAGC 149  
|||||  
34 rGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50  
|||||  
150 GGTAAAGATGGTGTGAAAACTTTTCATGACATGGACGAGGACTCTGAA 199  
|||||  
51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67  
|||||  
200 GAGAAAGAACTTTATCTAAACCTAGCTTTACATCTTGTCTGACTT 249  
|||||  
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84  
|||||  
250 CTTCCTCAAGCATCCTCGATAAAGATGTCGTTTACTGGTGGCTGTGTC 299  
|||||  
84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100  
|||||  
300 TTGCTGATATTTTCAGGATTTATGTCCTGAGGCTCTTACACGTCTCC 349  
|||||  
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117  
|||||  
350 GATAAATAAGGATATATTTATGTTTATAACAAGCGCACTAAGGGACT 399  
|||||  
117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu 134  
|||||  
400 AGAAGATACAAGAGCGCTCAATTAATAGGATATTTTATTACTTTGAGA 449  
|||||  
134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150  
|||||  
450 ACATTCATGGGTAAATCATATAACATATGCTTTGAGTTAGAGATAGC 499  
|||||  
151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167

|||||  
500 AATGAATCTTTACTCAACTTTTACAGAACATTATCTCAGTTATAACAA 549  
|||||  
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184  
|||||  
550 TGGCCACAATCAGAAAGTTTCATATGCACATGGTGGACCTCATGAGCTCTA 599  
|||||  
184 leIleCys 186  
|||||  
600 TCATTGT 607  
seq\_name: gb\_est43:AW175091

seq\_documentation\_block:  
LOCUS AW175091 703 bp mRNA EST 16-NOV-1999  
DEFINITION f132b08.y1 Sugano kawakami zebrafish DRB Danio rerio cDNA clone  
2639319 5' similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN  
; mRNA sequence.  
ACCESSION AW175091  
VERSION AW175091.1 GI:64411128  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Rasbora; Danio.  
REFERENCE  
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,  
Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,  
Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter  
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.  
TITLE WashU Zebrafish EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: S.L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA  
Sequencing by: Washington University Genome Sequencing Center  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 492.

FEATURES  
source  
1..703  
/organism="Danio rerio"  
/strain="AB"  
/db\_xref="taxon:7955"  
/clone="2639319"  
/clone\_lib="Sugano kawakami zebrafish DRB"  
/sex="mixed (one male and one female, including  
unfertilized eggs)"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage resistant)"  
/note="vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG);  
Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed  
with an oligo(dT) primer [ATGGGCTTTTCTTTTCTTTT];  
double-stranded cDNA was ligated to a DraIII adaptor  
[TGTGGCCCTACTGG], digested and cloned into distinct DraIII  
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site  
CACCATGTG). XhoI should be used to isolate the cDNA  
insert. Size selection was performed to exclude fragments  
<1.5kb. Library constructed by Dr. Sumio Sugano  
(University of Tokyo Institute of Medical Science) and  
kindly donated by Dr. Koichi Kawakami. Custom primers for  
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end  
primer CGACCTGCAGCTCGACACA."  
BASE COUNT 216 a 156 c 157 g 173 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 953.00 Length: 234



REFERENCE 1 (bases 1 to 683)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-femail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L14M8780 row: i column: 03  
 High quality sequence stop: 607.  
 Location/Qualifiers  
 1 . 683  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3598922"  
 /clone\_lib="NCI\_CGAP\_Mam6"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies, Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 235 a 146 c 138 g 163 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 974.50 Length: 226  
 Ratio: 4.575 Gaps: 4  
 Percent Similarity: 94.248 Percent Identity: 89.823

alignment\_block:  
 US-09-512-581-2 x BE532986 ..

Align seg 1/1 to: BE532986 from: 1 to: 583

978 AlaValSerGluLysLeu. LeuSerLeuLeuProGluTyrValValProt 994  
 |||||  
 3 GCTGTTAGTGAATAATATGTCTCTCTACCAGAGTATGTTGTCAT 52

994 yThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAsp 1010  
 |||||  
 53 ATACAATTACACCTTTGGCAGCATGCCAGATTATGTCAAGTACAGGAT 102

1011 IleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluI 1027  
 |||||  
 103 ATTGAAACAACTTAAGATGTGAAGAATGCCCTTTGTTGTTCTGGAGAT 152

1027 eLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysMetV 1044  
 |||||  
 153 ATTGATGGCTAAATAAGAAACCAACAGCCATCATTTATCAGAAAAATGG 202

1044 aGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAlaLys 1060  
 |||||  
 203 TAGAAAAATTAAACAGACAAAAGATGCTCAAGGACCATGATGATACAAA 252

1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleTle 1077  
 |||||  
 253 ATGAATGAAAAATTGTACACCTGTGTGATGTGTCATGGAACATCATCAT 302

1077 tSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeup 1094  
 |||||  
 303 GTTCAAGAGACACCATGACAGCTGGAGTCTCTTAAGGACCCCGTCTGC 352

1094 roAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsn 1110

|||||  
 353 CAGCTCGGTTTTTTCACCCAGCCTGACAAGAAATTTTAGTAACACCAAAAT 402

1111 TyrLeuProGluMetLysSerPhePheThrProGlyLysProLysTh 1127  
 |||||  
 403 TACCTGCTCCAGAAATGAATCATTTTCTACTCTCTGGAAAAACCTTAAAC 452

1127 rThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysG 1144  
 |||||  
 453 AGCCAATGTTCTCGAGCAGTAAATAAGCCACTTTTCATCAGCAGCAAAAC 502

1144 InSerGlnThrLysSerArgMetGluThrValSerAsnAla\_SerSe 1160  
 |||||  
 503 AGTCTCAGACCAAAATCATCAAGAATGGAACACTGTGAGCAACCAANGCAG 552

1160 rSerSerAsn\_ProSerSerProGlyArgIleLysGlyArgLeuAspSer 1176  
 |||||  
 553 CAGCTCCAAACCCCAAGCTCTCTGGAGGATCAGGGAGGCTTGATAGC 602

1177 SerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLe 1193  
 |||||  
 603 TCTGAATGGT.CACAGTGAACATGAAGATATACATGTC...TTCCCTTT 648

1193 uProGlyLysLysSerAspLys 1200  
 |||||  
 649 CCGCGAAAAAATGACAGAGAGA 670

seq\_name: gb\_est88:BF471938

seq\_documentation\_block:  
 LOCUS BF471938 608 bp mRNA EST 04-DEC-2000  
 DEFINITION UI-M-BH3-awf-g-03-0-UI.r1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-awf-g-03-0-UI 5', mRNA sequence.

ACCESSION BF471938  
 VERSION BF471938.1 GI:11541121

KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 608)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEst@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone Distribution:  
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
 should be noted that Bento Soares is generating a small number of  
 additional specialized non-redundant arrays of BMAP cDNAs whose  
 availability will be considered under appropriate and limited  
 collaborative arrangements  
 Seq primer: M13 Reverse.

FEATURES  
source

Location/Qualifiers  
 1 . 608  
 /organism="Mus musculus"  
 /db\_xref="C57BL/6J"  
 /strain="taxon:10090"  
 /clone="UI-M-BH3-awf-g-03-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged



403 CTAATAATGAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAT 452  
 1047 ilelysglnThrLysAspAlaGlnGlyProAspAlaLysMetAsnG1 1063  
 453 ATTAACAACAAGAGATGCCAGGACCATGATGCAAAATGAATGA 502  
 1063 uLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLys 1080  
 503 AAAACTGACATGCTGTGATGTCATGTCATGATATCATCATCATCAAGA 552  
 1080 erThrThrTyrSerLeuGluSerProLysAspProValLeuPro 1094  
 553 GTACTACATACAGTTTGAATCTCTTAAGACCCGGTACTACCA 596

seq\_name: gb\_est100:BG387646

seq\_documentation\_block:  
 LOCUS BG387646 802 bp mRNA EST 12-MAR-2001  
 DEFINITION 602412496F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4521095 5',  
 mRNA sequence.

ACCESSION BG387646  
 VERSION BG387646.1 GI:13281092  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 802)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLAM10419 row: p column: 24  
 High quality sequence stop: 637.

## FEATURES

source

1..802  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4521095"  
 /clone\_lib="NIH\_MGC\_92"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; Oligo-dr primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."  
 BASE COUNT 270 a 149 c 167 g 216 t  
 ORIGIN

## alignment\_scores:

Quality: 1023.00 Length: 256  
 Ratio: 4.316 Gaps: 7  
 Percent Similarity: 92.578 Percent Identity: 87.500

## alignment\_block:

US-09-512-581-2 x BG387646

Align seg 1/1 to: BG387646 from: 1 to: 802

365 AlaIleArgHisValIleValSerIleValThrAlaAlaLysLysAs 381  
 1 GCTATTAGACATGATGTTATGTGTCATAATGTTACAGCTGCTCAAAAGGA 50

381 pIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrL 398  
 51 TATTCCTCTGTCATCATCATCTACTTAATTTTGTGAGAGAGAGAACAT 100  
 398 euAspLysArgTTPArgValArgLysGluAlaMetMetGlyLeuAlaGln 414  
 101 TAGACAAACGATGGAGAGTACGCAAGAACCCATGATGGGACTTGCCAA 150  
 415 IleTyrLysLysTyrAlaLeuGlnSer..AlaAlaGlyLysAspAlaAla 431  
 151 ATTTATAAGAAATATGCTTTACAGTCAGACGCTGGAAAAGATGCTGCAA 200  
 431 ysglnIleAlaTTPileLysAspLysLeuLeuHisIleTyrTyrGlnAsn 447  
 201 AACAGATAGCATGGATCAAGACAAATTTGTACATATATATATCAAAAT 250  
 448 SerIleAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVa 464  
 251 AGTATTGATGATCGACTACTTGTGAACGGATCTTGTCTCAATACATGGT 300  
 464 lProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuT 481  
 301 TCCTCAATTTAGAACTACAGAACGGATGAAATGCTTATATTTACTTGT 350  
 481 YrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTTPLys 497  
 351 ATGCCACATCGGATTAAATGCTGTGNAAGCATTTGAATGACATGTGAAA 400  
 498 CysGlnAsnLeuLeuArgHisGlnVal.LysAspLeuLeuAspLeuIleL 514  
 401 TGTCAAAATCTGCTCCGACATCAAGTACAAGGATTTGCTTGACTTGATTA 450  
 514 ysgln..ProLysThrAspAlaSerValLysAlaIlePheSer..LysValM 530  
 451 AGCAAAACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAAGTGA 500  
 530 etValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546  
 501 TGGTTATTACAAGAAATTTACCTGATCTCTGTAAGGCTCAGGATTTTCATG 550  
 547 LysLysPheThrGlnValLeuGluAsp.....AspGluLysIleArgLy 561  
 551 AAGAAATTCACACAGAGTGTTCAGACAGATGACTGAGGACAAATCAAGAAA 600  
 561 sGlnLeuGlu..ValLeuValSerProThrCysSerCys..LysGlnAlaG 577  
 601 GCAGTTTAGAAAGTACTTGTGTAGTCCAAACATGCTCTCTGCCAAAGCGAGCT 650  
 577 luGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPro 593  
 651 GACAGTGTGTGCGTGACATAACCTAGAGTTGGCAACCCCAACACAGACT 700  
 594 ThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPr 610  
 701 ACAAGATC..TTCCTGGCAATGCTCACGCTTCTCTTGGAGACGATAGCTCT 749  
 610 ovalHis 612  
 750 GTGCCAT 756

seq\_name: gb\_est75:BE532986

seq\_documentation\_block:  
 LOCUS BE532986 683 bp mRNA EST 09-AUG-2000  
 DEFINITION 601235013F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3598922 5',  
 mRNA sequence.  
 ACCESSION BE532986  
 VERSION BE532986.1 GI:9761631  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.



201 AGAAGTATCTCTGACACAAATCGTCAAAATTCAGGCTATTAATAATGATGG 250  
833 alArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849  
|||||  
251 TTTCGATGGCTACTTGAATGAAAAATAATACAGTAAATCAGCAACTTCT 300  
850 ThrLeuArgLeuLeuThrThrLeuLeuHisSerAspGlyAspLeuThrG1 866  
|||||  
301 ACCTTAAGATTCTTAACAAATAATGTCATAGTGTGAGAGCTTGACAGA 350  
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaG 883  
|||||  
351 ACAGGGGAAATTAAGTAACACAGATATGTCAGCTCTGAGACTTGCTG 400  
883 lYserAlaIleValLysLeuAlaGlnGluProCysTyrHisGluLeuIle 899  
|||||  
401 CGAGTGCTATTGTAAGCTGGCACAAGAACCCCTGTACCATCAATCATC 450  
900 ThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysty 916  
|||||  
451 ACATTAGAACAAATACAGTATGTCATTAGCTATCAACGATGAATGCTA 500  
916 rGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgL 933  
|||||  
501 TCAAGTAGACAAAGTGTGTCGCCAGAACTTCACAAAGCCCTTCCCGTT 550  
933 euArg. LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLys. A 949  
|||||  
551 TACCGGCTTCCACTTGAGTATATGCAATCTGTGCCCTTGTGGTCAAGAAG 600  
949 spProValLys...GluArgArgAlaHisAlaArgGlnCysLeuValLys 964  
|||||  
601 ATCCCTGTAAAGGAGAAGAGAGCTCATGCTAGGCAATGTTTGTGGTGAAC 650  
965 AsnIleAsn.ValArgArgGlu.TyrLeuLysGlnHisAlaAlaValSer 980  
|||||  
651 ATATAAATGTGACGGGGAGCTATCTTGAAGCAGATGGGCTGTAGT 700  
981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHi 997  
|||||  
701 GAGAACTTATGGTCTCTTACACAGAT.TTGGTGCCCTATATGATTCA 749  
997 sLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnL 1014  
|||||  
750 CTTTGGGCGCTTGGCCCC.....ATTATTAAAGCGGGATTGAGCACT 793

seq\_name: gb\_est75:BE504550

seq\_documentation\_block: 596 bp mRNA EST 04-AUG-2000  
LOCUS BE504550  
DEFINITION h258a05.x1 NCI-CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3212144 3'  
similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN ; , mRNA  
sequence.  
ACCESSION BE504550  
VERSION BE504550.1 GI:9706958  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -400P from Gibco  
High quality sequence stop: 458.  
Location/Qualifiers  
source  
1. 596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3212144"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI-CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 211 a 103 c 117 g 165 t  
ORIGIN

alignment\_scores:

Quality: 1027.00 Length: 198  
Ratio: 5.187 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-512-581-2 x BE504550

Align seg 1/1 to: BE504550 from: 1 to: 596

897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAs 913  
|||||  
3 GAAATCATCATATTAGAACAAATATCAGCTATGTCATAGCTATCAACGA 52  
913 pGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyL 930  
|||||  
53 TGAATGCTATCAAGTAAGACAAGTGTGTCGCCAGAACTTCACAAAGGCC 102  
930 euSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946  
|||||  
103 TTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGT 152  
947 AlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuVa 963  
|||||  
153 GCAAAAGATCCTGTGAAGAGAGAGAGAGCTCATGCTAGCAATGTTGCT 202  
963 lYsAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValS 980  
|||||  
203 GAAAAATATAAATGTAAAGCGGGAGTATCTGAAGCAGCATGCAGCTGTA 252  
980 erGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996  
|||||  
253 GGGAAAAATATTGTCTCTCTACAGAGATGTTGTTCCATATACAAT 302  
997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluG1 1013  
|||||  
303 CACCTTTGGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACA 352  
1013 nLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMeta 1030  
|||||  
353 ACTTAAAGATGTTAAAGAATGCTTTGTTGTTCTCTGGAATATTAAATGG 402  
1030 laLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046  
|||||



```

53  TTCCACCCCTCAGATTACTTAACAACGATAGTCATAGTGGGATTGA 102
865  hrGluGlnGlyLysIleSerLysProAspMetSerArgLeuAla 881
|||||
103  CAGAACAGGAAATAGTAACACAGATATGTCACGCTGAGACTGCT 152
|||||
882  AlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 898
|||||
153  GCTGGGAGTGCTATTGTGAAGCTGGCACAGGACCCCTGTACCACGAGAT 202
|||||
898  eileThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluC 915
|||||
203  CATTAACACTGGACGACGATACCAGCTGTGTGATAGCATCAATGATGAGT 252
|||||
915  ystYrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931
|||||
253  GTTATCAAGTCAGGCGAGGTGTGCTGCAGAACTTCACAAAGCCCTTCC 302
|||||
932  ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla 948
|||||
303  CGCTTACGGCTTCCCTTGAGTACATGGCCATCTGTCTTTGTGCCAA 352
|||||
948  sAspProValLysGluArgAlaHisAlaArgGlnCysLeuValLys 965
|||||
353  AGACCCCTGTGAAAGAGAGCGGACCCATGCTAGACAGTGTCTGGTGAAGA 402
|||||
965  snIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu 981
|||||
403  ACATCACTGTGAGGAGGAGTACTCTGAAGCAGCATGCAGCTGTAGTGA 452
|||||
982  LysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLe 998
|||||
453  AAATTATTGCTCTTCTACCAGAGTATGTGGTTCATATACATTCACCT 502
|||||
998  uLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeu 1015
|||||
503  TTGGGCACATGACCCAGATATGATCAACGTACAGGATATTTGAACAACTTA 552
|||||
1015  ysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLys 1031
|||||
553  AAGATGTGACAAAGTAATCCCTGGTTT.GTCTGGAGATATTGATGGGTAA 601
|||||
1032  AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLy 1048
|||||
602  AATGAAACAAACAG.CATGCTTTATCAGAAATGGTAGAAATATTAA 650
|||||
1048  sGlnThrLysAspAlaGlnGlyProAspAlaLysMetAsnGluLysL 1065
|||||
651  CAGACCAAGATTTGCTCCAGGACCGATGAT.....CACCACCTGAC 691
|||||
1065  euTyrThr.....ValCysAspValAlaMetAsnIleIleMetSerLys 1079
|||||
692  TGCACAACTCTCCGCTGTGTGTGTTGCCATGGACTCATCAGTTCAGAGA 741
|||||
1080  SerThrTyrSerLeuGluSerPro..... 1088
|||||
742  ...ACACGTACAGCTGGAGTTCTAGACCCGGTCCGCTGCTTTAAACA 788
|||||
1089  .....LysAspProValLeuProAlaArgPhe.....T 1099
|||||
789  GCGGAGAAATAGTACCAATTCGCTCCAATGATATTTACTCTGGACTAA 838
|||||
1099  hrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu..... 1112
|||||
839  GCAGTCTCGACAAAACTTT.....CGGACCTTCACCTAATGATGC 879
|||||
1113  .....ProGluMetLysSerPheThrProGlyLysProLysTh 1127
|||||
880  ACGCCGCCCGGAGGAGGCGACACGCGACACCCCTCGACAACTATCCA 929
|||||
1127  rThrAsn 1129
|||||
930  CACATCT 936

```

seq\_name: gb\_est80:BE873840

seq\_documentation\_block: BE873840 mrna EST 20-OCT-2000  
 LOCUS 601483992F1 NIH\_MGC\_69 Homo sapiens cDNA clone IMAGE:3886511 5',  
 mrna sequence.  
 ACCESSION BE873840  
 VERSION BE873840.1 GI:10322616  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 848)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLAM9663 row: g column: 24  
 High quality sequence stop: 654.  
 Location/Qualifiers

FEATURES  
 source

1..848  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3886511"  
 /clone\_lib="NIH\_MGC\_69"  
 /tissue\_type="large cell carcinoma, undifferentiated"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."

BASE COUNT 261 a 163 c 191 g 232 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 1061.50 Length: 271  
 Ratio: 4.298 Gaps: 6  
 Percent Similarity: 91.144 Percent Identity: 86.347

alignment\_block:  
 US-09-512-581-2 x BE873840 ..

Align seg 1/1 to: BE873840 from: 1 to: 848

750 ProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrPr 766  
 |||||  
 2 CCTCTCATGAAGAGCTTAGATCCAAACACCTGGAACATCTCATAACACC 51  
 |||||  
 766 oLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAla 783  
 |||||  
 52 ATTGTTTACTAT.GGTCAATATGCTCTCTTCCTTGACCTGATCAATTGCTG 100  
 |||||  
 783 laProTyrLysSerTrpValAlaThrPheIleValLysAspLeuMet 799  
 |||||  
 101 CTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATG 150  
 |||||  
 800 AsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProaspG 816  
 |||||  
 151 AATGATCGGCTTCAGGGAAGAAACAACTAAACTTTGGGTTCAGATGA 200  
 |||||  
 816 uGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetV 833  
 |||||



/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCATAGCGCCGAGCGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 225 a 129 c 144 g 231 t  
ORIGIN

alignment\_scores:  
Quality: 1091.00 Length: 242  
Ratio: 4.662 Gaps: 7  
Percent Similarity: 96.694 Percent Identity: 95.868

alignment\_block:

US-09-512-581-2 x BG432960 ..

Align seg 1/1 to: BG432960 from: 1 to: 729

122 SerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVa 138  
2 AGCCACAGTTCATAGGTATTTTATTTACTTGAGACAT\_GCTTGGGT 50  
138 llySerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheT 155  
51 CAAGTCATATACATATGCTTTGAGTTAGATAGCATGCAATGAATTTTCA 100  
155 hrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGln 171  
101 CCCAGCTATACAGAACCTTATTTTTCAGTTTATAACAATGGCCACATCAG 150  
172 LysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluG1 188  
151 AAAGTCCATATGCATGCTAGACCTTATGAGCTCTATTATTGTTGAAGG 200  
188 yAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnValP 205  
201 TGATACAGTCTCTCAGGAGCTTTTGGATACGGTTTTTAGTAAATCTGGTAC 250  
205 roAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu 221  
251 CTGCTCATAGAATTTAAACAAGCAAGCATATGATTGGCAAGGCTTTTA 300  
222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAs 238  
301 CTGAAGAGGACAGCTCAAGCTATTGAGCCATATATACCAATTTTTTAA 350  
238 nGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisV 255  
351 TCAGGTTCTGATGCTTGGGAAACATCTATACGCGATTTGTCAGAGCATG 400  
255 al.PheAspLeuLeuGluLeuTyrAsnIleAspSerHisLeuLeuLe 271  
401 TCTTTTGACTTAATTTGGAGCTCTACAATATTGATGATTCATTTGCTGCT 450  
271 uSerValLeuProGlnLeu.GluPheLysLeuLysSerAsnAspAsnGlu 287  
451 CTCTGTTTACCCCGCTTGGATTTAAATTAAGAGCAATGATATAGAG 500  
288 GluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAs 304  
501 GAGCCCTACAGTTGTTAACTACTGCGAAAAATGTTTGGGCGCAAGGA 550  
304 pSerGluLeuAlaSerGln.AsnLysProLeuTrpGlnCysTyr.LeuG1 320  
551 TTCAGAAATGGCTTCTCAAAAACAAGCCACTTTGGCAGTGCTACTTTGGG 600  
320 yArgPheAsnAspIleHisValProIle.ArgLeuGluCysVal.LysPh 336

|||||  
601 CAGGTTTATGATATCCATGTACCAATTCGCGTGGATGTGTGACCAT 650  
336 eaLaserHisCysLeuMetAsnHisProAspLeuAla.LysAspLeuThr 352  
651 TCTAGCCATTGCTCATGAACCATCTGATTAGCAAAAGATTTAACA 700  
353 GluTyrLeuLys 356  
701 GAGTATCTTAAA 712  
seq\_name: gb\_est84:BF142489

seq\_documentation\_block:

LOCUS BF142489 944 bp mRNA EST 24-OCT-2000  
DEFINITION 601786573F1 NCI\_CGAP\_Lu30 Mus musculus cDNA clone IMAGE:4016155 5', mRNA sequence.  
ACCESSION BF142489  
VERSION BF142489.1 GI:10981529  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 944)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLAM9263 row: a column: 20  
High quality sequence stop: 620.  
FEATURES  
Location/Qualifiers  
1..944  
/organism="Mus musculus"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4016155"  
/clone\_lib="NCI\_CGAP\_Lu30"  
/tissue\_type="tumor, metastatic to mammary"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; transgenic model WNT-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies.  
Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 277 a 221 c 218 g 228 t  
ORIGIN

alignment\_scores:

Quality: 1066.00 Length: 319  
Ratio: 4.069 Gaps: 7  
Percent Similarity: 82.132 Percent Identity: 74.295

alignment\_block:

US-09-512-581-2 x BF142489 ..

Align seg 1/1 to: BF142489 from: 1 to: 944

832 MetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyTh 848  
3 ATGGTTTCGATGGCTACTTGGATGAAGAATAATACAGTAAGTCAGGAAC 52  
848 rSerThrLeuArgLeuLeuThrIleLeuHisSerAspGlyAspLeuT 865  
|||||



```

DEFINITION 602370939F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4479153 5',
            mRNA sequence.
ACCESSION  BG256731
VERSION    BG256731.1  GI:12766547
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 734)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10310 row: m column: 10
            High quality sequence stop: 732.
            Location/Qualifiers
                1..734
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4479153"
                /clone_lib="NIH_MGC_92"
                /tissue_type="embryonal carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT  278 a 154 c 161 g 141 t
ORIGIN
alignment_scores:
    Quality: 1143.50      Length: 241
    Ratio: 4.887          Gaps: 1
    Percent Similarity: 97.095 Percent Identity: 95.021
alignment_block:
US-09-512-581-2 x BG256731
Align seg 1/1 to: BG256731 from: 1 to: 734
1078 SerLysSerThrThrThrSerLeuGluSerProLysAspProValLeuPr 1094
1 TCAAAGAGTACTACATACAG.TTGGATCTCTCTAAAGACCCGGTACTACC 49
1094 oAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnT 1111
50 AGCTCGTTCTTCTCACCACTGACACAGAAATTCAGTAACACCAAAATTT 99
1111 yrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThr 1127
100 ATCTGCCTCTGAAATGAAATCATCTTTTCTACTCTCTGAAAAACCTAAACA 149
1128 ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGI 1144
150 ACCAATGTTCTAGAGCTGTGTAAACAGCCACTTTCATCAGCGCAAGCA 199
1144 nSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSers 1161
200 ATCTCAGACCAATCATCAGCATGGAATGGAATGTAAAGCAATGCAGCAGCA 249
1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSer 1177

```

```

250 GCTCAATCCAAAGCTCTCTCTGGAGAAATAAAGGGAGGCTTGATAGTTCT 299
1178 GluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuPr 1194
300 GAAATGGATCACAGTGAATGAAGATTACACAATGCTTCACCTTTGCC 349
1194 oGlyLysLysSerAspLysArgAspSerAspSerLeuValArgSerGluL 1211
350 GGGGAAAAAAGTGCACAGAGACGACCTCTGATCTTGTAAGGCTCGAAT 399
1211 euGluLysProArgGlyArgLysLysThrProValThrGluGlnGluGlu 1227
400 TGGAGAGGCTTAGAGCAGGAAAAAAGCCCGCTCACAGAACAGGAGGAG 449
1228 LysLeuGlyMetAspLysLeuThrLysLeuValGlnGlnGlnLysProLy 1244
450 AAATTAGGATGATGACTTGACTTAAGTTGGTACAGGAACAGAAACCTAA 499
1244 sGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1261
500 AGCAGTCAGCGAAGTCGGAAAGAGCCATACGGCTTCAGATCTGATG 549
1261 lucGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsn 1277
550 AACAGCAGTGGCGCTGAGGAAAAAGAGGCTCAAGAAGATATATTAGAACA 599
1278 ...GluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPr 1293
600 CTGAACGATGACAGCAATAGTCCGCCAAAAACAAGGTACAGAGCGCGAC 649
1293 oProLysProLeuGlyGlyThrProLysGluGluProThrMetLysT 1310
650 CACAAAAACCTCTTGGTGGAGGTACACCAAAAGAGAGCCAAACGATGAAA 699
1310 hrSerLysLysGlySerLysLys 1317
700-CTTCTATAAAGGAAGCAACAA 722
seq_name: gb_est100:BG432960
seq_documentation_block:
LOCUS      BG432960              729 bp      mRNA           EST
DEFINITION 602495783F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4609533 5',
            mRNA sequence.
ACCESSION  BG432960
VERSION    BG432960.1  GI:13339466
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 729)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1353 row: e column: 22
            High quality sequence stop: 722.
            Location/Qualifiers
                1..729
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4609533"
                /clone_lib="NIH_MGC_75"
                /lab_host="DH10B (T1 phage-resistant)"
FEATURES
source

```



g1\_gss22: A2185970 + 160.00 245.85 0.0001 886 ! AZ185970 SP\_1006\_AL\_F07\_T7A STD  
g1\_gss9: A0269542 + 160.00 245.67 0.0001 899 ! A0269542 HS\_2022\_AL\_H12\_T7 CIT  
g1\_est100: BG432918 + 160.00 244.78 0.0002 966 ! BG432918 602495738F1 NIH\_MGC\_75  
est110: BF631435 - 160.00 244.66 0.0002 975 ! BF631435 HVSM60015P03f Hordeum  
g1\_est76: BF612455 + 160.00 244.63 0.0002 978 ! BF612455 601451835F1 NIH\_MGC\_66  
g1\_est89: BF529509 + 160.00 244.41 0.0002 995 ! BF529509 602043272F1 NCI\_CGAP\_B  
g1\_est101: BF529840 + 160.00 244.07 0.0002 1023 ! BF529840 602358927F1 NIH\_MGC\_6  
g1\_est110: BF631286 - 160.00 243.33 0.0002 1086 ! BF631286 HVSM60015I03f Hordeum  
g1\_est98: BG288980 + 160.00 242.34 0.0002 1176 ! BG288980 602381239F1 NIH\_MGC\_9  
g1\_est89: BF573791 + 160.00 242.01 0.0002 1207 ! BF573791 602077718F1 NIH\_MGC\_6

seq\_name: gb\_est88: BF509252

seq\_documentation\_block:

LOCUS BF509252 791 bp mRNA EST 06-DEC-2000  
DEFINITION UI-H-B14-aow-c-07-0-UI-sl NCI\_CGAP\_Sub8 Homo sapiens cDNA clone  
IMAGE: 3086220 3', mRNA sequence.

ACCESSION BF509252

VERSION BF509252.1 GI: 11592550

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 791)

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:

NCI-CCGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

Location/Qualifiers

1..791

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone.lib="NCI\_CGAP\_Sub8"

/lab.host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; NCI\_CGAP\_Sub8

is a subtracted library derived from NCI\_CGAP\_Sub5. The

NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A

single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used

as a tracer in a subtractive hybridization with a driver

comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE

clone ids 2732833-2737415, 3068040-3069191; 25% of the

driver population), a pool of clones from NCI\_CGAP\_Sub4

(IMAGE clone ids 2723592-2729326; 25% of the driver

population), NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE IDs

2728969-2733190; 25% of the driver population), and

NCI\_CGAP\_Sub7 (IMAGE IDs 3069192-3072238, 3081864-3084550

; 25% of the driver population). Subtraction was

performed as previously described (Bonaldo, Lennon &

Soares (1996): Normalization and Subtraction: Two

Approaches to Facilitate Gene Discovery. Genome Research

6, 791-806.

TAG\_LIB=NCI\_CGAP\_Lu5

TAG\_TISSUE=lung

TAG\_SEQ=CAAC\*

BASE COUNT 270 a 151 c 144 g 226 t

ORIGIN

alignment\_scores:

seq\_name: gb\_est98: BG256731

seq\_documentation\_block:

LOCUS BG256731

734 bp mRNA

EST

13-FEB-2001



gb_est76:BE612723	162.00	246.67	0.0001	1090	BE612723	601452584F1	NIH_MGC_6	gb_est101:BG492841	161.00	245.02	0.0002	1086	BG492841	602536869F1	NIH_MGC
gb_est80:BE886879	162.00	246.61	0.0001	1095	BE886879	601507854F1	NIH_MGC_7	gb_est101:BG501134	161.00	244.72	0.0002	1113	BG501134	602546367F1	NIH_MGC
gb_est102:BG538263	162.00	246.56	0.0001	1100	BG538263	602566730F1	NIH_MGC_7	gb_est101:BG529642	161.00	244.59	0.0002	1124	BG529642	602558047F1	NIH_MGC
gb_est97:BG1177849	162.00	246.48	0.0001	1107	BG1177849	602314257F1	NIH_MGC_8	gb_est80:BE910153	161.00	243.22	0.0002	1256	BE910153	601503441F1	NIH_MGC
gb_est97:BG115045	162.00	246.38	0.0001	1120	BG115045	602315477F1	NIH_MGC_8	gb_est76:BE9113921	161.00	242.79	0.0002	1300	BE9113921	601504041F1	NIH_MGC
gb_est93:BF860254	162.00	245.29	0.0002	1218	BF860254	963016A10_Y2_C_reinh		gb_est86:BF342615	161.00	242.78	0.0002	1301	BF342615	602013750F1	NCI_CG
gb_est19:BF279708	162.00	245.21	0.0002	1226	BF279708	GA_EB0033F16_Gossyp		gb_est92:BF792674	161.00	242.77	0.0002	1302	BF792674	602238057F1	NIH_MGC
gb_est95:BF981306	162.00	245.03	0.0002	1244	BF981306	602308052F1	NIH_MGC_8	gb_est86:BF708736	161.00	242.19	0.0002	1364	BF308736	601890033F1	NIH_MGC
gb_est96:BG107670	162.00	243.52	0.0002	1401	BG107670	602277836F1	NIH_MGC_8	gb_est83:BF104114	161.00	241.65	0.0002	1425	BF104114	601647715F1	NIH_MGC
gb_est91:BG159486	162.00	243.78	0.0002	1491	BG159486	602577760F1	NCI_CGAP	gb_est80:BE983934	161.00	241.36	0.0003	1459	BE983934	601437711F1	NIH_MGC
gb_est95:BG025954	162.00	242.52	0.0002	1523	BG25954	602292113F1	NIH_MGC_8	gb_est89:BF576555	161.00	240.54	0.0003	1558	BF576555	602143993F1	NIH_MGC
gb_est83:BF131624	162.00	242.21	0.0002	1562	BF131624	601820432F1	NIH_MGC_8	gb_est100:BG390268	161.00	240.50	0.0003	1563	BG390268	602416232F1	NIH_MGC
gb_est85:BF219450	162.00	242.14	0.0002	1570	BF219450	601884243F1	NIH_MGC_8	gb_est98:BG356684	161.00	240.22	0.0003	1599	BG356684	602370873F1	NIH_MGC
gb_est97:BG113909	162.00	242.12	0.0002	1574	BG113909	602284367F1	NIH_MGC_8	gb_est82:BF038463	161.00	239.97	0.0003	1632	BF038463	601460525F1	NIH_MGC
gb_est95:BG031363	162.00	242.11	0.0002	1573	BG031363	602290655F1	NIH_MGC_8	gb_est98:BG388709	161.00	239.83	0.0003	1639	BG388709	602385523F1	NIH_MGC
gb_est93:BF868150	162.00	241.59	0.0002	1642	BF868150	963100G03_Y1_C_reinh		gb_est86:BF344711	161.00	239.70	0.0003	1668	BF344711	602017413F1	NCI_CG
gb_est84:BF204687	162.00	241.53	0.0002	1650	BF204687	601867525F1	NIH_MGC_1	gb_est92:BF797714	161.00	239.63	0.0003	1677	BF797714	602025765F1	NIH_MGC
gb_est100:BG394047	162.00	241.25	0.0003	1687	BG394047	602456205F1	NIH_MGC_1	gb_est97:BG114716	161.00	239.61	0.0003	1680	BG114716	602315386F1	NIH_MGC
gb_est84:BF180004	162.00	240.83	0.0003	1745	BF180004	601806351F1	NCI_CGAP	gb_est97:BG121828	161.00	239.59	0.0003	1683	BF121828	602350823F1	NIH_MGC
gb_est83:BF122002	162.00	240.65	0.0003	1771	BF122002	601756344F1	NCI_CGAP	gb_est91:BF676264	161.00	239.53	0.0003	1691	BF676264	602308961F1	NIH_MGC
gb_est93:BF124728	162.00	240.65	0.0003	1771	BF124728	601761407F1	NCI_CGAP	gb_est92:BF789291	161.00	239.24	0.0003	1732	BF789291	602104707F1	NCI_CG
gb_est93:BF121855	162.00	240.32	0.0003	1819	BG112855	602285053F1	NIH_MGC_8	gb_est83:BF099591	161.00	239.03	0.0003	1759	BF099591	601751161F1	NCI_CG
gb_est85:BF302581	162.00	240.17	0.0003	1841	BF302581	602032428F1	NCI_CGAP	gb_est84:BF166246	161.00	238.90	0.0003	1778	BF166246	601777007F1	NCI_CG
gb_est94:BF905119	162.00	240.16	0.0003	1842	BF965119	602268893F1	NIH_MGC_8	gb_est89:BF539718	161.00	238.78	0.0003	1796	BF539718	602049929F1	NCI_CG
gb_est76:BE612548	162.00	240.07	0.0003	1856	BF612548	601452606F1	NIH_MGC_8	gb_est95:BG333593	161.00	238.35	0.0004	1859	BG333593	602301177F1	NIH_MGC
gb_est97:BF115862	162.00	239.51	0.0003	1942	BG115862	602168842F1	NIH_MGC_8	gb_est91:BF673971	161.00	238.05	0.0004	1939	BF673971	602137356F1	NIH_MGC
gb_est94:BF144133	162.00	239.27	0.0003	1979	BF144133	602188425F1	NCI_CGAP	gb_est98:BG244238	161.00	237.83	0.0004	1939	BG244238	602356005F1	NCI_CG
gb_est94:BF167018	162.00	239.12	0.0003	2004	BF67018	60174161F1	NCI_CGAP	gb_est89:BF570927	161.00	237.77	0.0004	1948	BF70927	602076125F1	NIH_MGC
gb_est91:BF673999	162.00	237.33	0.0004	2315	BF673999	602137389F1	NIH_MGC_8	gb_gss10:AV761512	160.50	249.07	9.3e-05	732	AQ761512	nbsb0020424B1	CUG_R
gb_gss16:AQ743461	161.50	248.52	0.0001	877	AQ743461	HS_5482_B1_A01_SP6_RPC		gb_est41:AV761512	160.50	248.93	9.5e-05	740	AV761512	av761512_MDS_Homb	R
gb_gss22:AZ200698	161.50	248.47	0.0001	891	AZ200698	SP_1026_A1_A04_SP6E_SP		gb_est101:BG499298	160.50	248.28	0.0001	780	BG499298	602545761F1	NIH_MGC
gb_est101:BG493864	161.50	248.33	0.0001	891	BG493864	60253874F1	NIH_MGC_59	gb_est101:BG433046	160.50	246.76	0.0001	882	BG433046	602017163F1	NCI_GA
gb_est93:BF864215	161.50	247.66	0.0001	901	BF864215	963050A04_x1_C_reinh		gb_gss30:AZ670461	160.50	246.74	0.0001	883	AZ670461	ENTKY42TF_Entamoeba	
gb_est98:BG246290	161.50	247.30	0.0001	968	BG246290	602360087F1	NCI_CGAP_M	gb_est101:BG529491	160.50	246.11	0.0001	929	BG529491	602557839F1	NIH_MGC
gb_est79:BE779116	161.50	247.03	0.0001	989	BE779116	601464725F1	NIH_MGC_67	gb_est102:BG542951	160.50	245.30	0.0002	992	BG542951	602569345F1	NIH_MGC
gb_est101:BG501043	161.50	246.89	0.0001	1000	BG501043	602548450F1	NIH_MGC_8	gb_est100:BG391723	160.50	244.16	0.0002	1049	BG391723	602418010F1	NIH_MGC
gb_est98:BG252634	161.50	246.68	0.0001	1007	BG252634	602366272F1	NIH_MGC_8	gb_est96:BG107887	160.50	244.59	0.0002	1050	BG107887	602278155F1	NIH_MGC
gb_est98:BG291331	161.50	246.38	0.0001	1042	BG291331	602388495F1	NIH_MGC_8	gb_est83:BF1119947	160.50	244.13	0.0002	1090	BF119947	6017580229F1	NCI_CG
gb_gss2:CN50165N	161.50	246.35	0.0001	1045	AL106325	Drosophila_melanogast		gb_est100:BG3545095	160.50	243.57	0.0002	1140	BG3545095	602507103F1	NIH_MGC
gb_est102:BG545612	161.50	246.23	0.0001	1055	BG545612	602572869F1	NIH_MGC_7	gb_est100:BG391723	160.50	243.55	0.0002	1142	BG391723	602418010F1	NIH_MGC
gb_est89:BF531096	161.50	246.12	0.0001	1064	BF531096	602072257F1	NCI_CGAP	gb_est93:BF865614	160.50	243.44	0.0002	1152	BF865614	963061A05_Y1_C_reinh	
gb_est101:BG479714	161.50	244.76	0.0002	1188	BF79714	602527183F1	NIH_MGC_2	gb_est100:BG426367	160.50	243.26	0.0002	1169	BG426367	602509021F1	NCI_CG
gb_est89:BF530242	161.50	243.76	0.0002	1287	BF530242	602071417F1	NCI_CGAP	gb_gss14:AQ621890	160.50	243.24	0.0002	1171	AQ621890	HS_3107_B1_G12_R	
gb_est87:BF381856	161.50	243.25	0.0002	1341	BF381856	601816123F1	NIH_MGC_5	gb_gss16:AQ6780342	160.50	243.20	0.0002	1194	AQ6780342	HS_3184_A2_D07_77C	
gb_est86:BF315178	161.50	242.48	0.0002	1427	BF315178	601899541F1	NIH_MGC_1	gb_est100:BG433046	160.50	242.93	0.0002	1200	BG433046	602507102F1	NIH_MGC
gb_est92:BF032144	161.50	242.47	0.0002	1438	BF032144	601595575F1	NIH_MGC_5	gb_est95:BG034865	160.50	242.89	0.0002	1204	BG034865	602325322F1	NIH_MGC
gb_est86:BF309252	161.50	242.06	0.0002	1476	BF309252	601887702F1	NIH_MGC_1	gb_est97:BE784623	160.50	242.56	0.0002	1237	BE784623	601474059F1	NIH_MGC
gb_est97:BG115872	161.50	241.99	0.0002	1485	BG115872	602316863F1	NIH_MGC_8	gb_est81:BE965081	160.50	241.79	0.0002	1316	BE965081	601638689F1	NIH_MGC
gb_est80:BE74969	161.50	241.97	0.0002	1487	BE74969	601487528F1	NIH_MGC_6	gb_est97:BF180352	160.50	241.50	0.0002	1347	BF180352	602331221F1	NIH_MGC
gb_est93:BF863301	161.50	241.84	0.0002	1503	BF863301	963042F06_Y1_C_reinh		gb_est89:BF533898	160.50	241.14	0.0003	1387	BF533898	602075261F1	NCI_CG
gb_est79:BE754102	161.50	241.29	0.0003	1571	BF754102	601754666F1	NCI_CGAP	gb_est89:BF574012	160.50	240.97	0.0003	1406	BF574012	602131843F1	NIH_MGC
gb_est85:BF257135	161.50	241.20	0.0003	1582	BF257135	602370720F1	NIH_MGC_9	gb_est98:BG249529	160.50	240.84	0.0003	1421	BG249529	602319586F1	NIH_MGC
gb_est100:BG387830	161.50	241.05	0.0003	1601	BF214528	601845791F1	NIH_MGC_5	gb_est97:BG166139	160.50	240.81	0.0003	1424	BF166139	602340748F1	NIH_MGC
gb_est97:BG121860	161.50	240.37	0.0003	1632	BG387830	602412732F1	NIH_MGC_5	gb_est100:BG421340	160.50	240.40	0.0003	1472	BG421340	602451464F1	NIH_MGC
gb_est92:BF782966	161.50	239.67	0.0003	1790	BG121860	602309070F1	NIH_MGC_9	gb_est83:BF118725	160.50	240.22	0.0003	1494	BF118725	601755125F1	NCI_CG
gb_est100:BG389220	161.50	239.46	0.0003	1821	BG389220	602413855F1	NIH_MGC_9	gb_est97:BF95690	160.50	240.20	0.0003	1496	BF95690	601590686F1	NIH_MGC
gb_est72:BE289217	161.50	239.24	0.0003	1853	BE289217	601100574F1	NCI_CGAP	gb_est101:BG482357	160.50	239.68	0.0003	1560	BG482357	602525612F1	NIH_MGC
gb_est84:BF195517	161.50	239.06	0.0004	1860	BF185517	601804606F1	NIH_MGC_5	gb_est83:BF687013	160.50	239.29	0.0003	1610	BF104133	601647740F1	NIH_MGC
gb_est86:BF339426	161.50	238.73	0.0004	1931	BF339426	602308773F1	NIH_MGC_5	gb_est91:BF687013	160.50	238.46	0.0004	1721	BF687013	601027290F1	NCI_CG
gb_gss16:AQ779894	161.00	249.65	8.7e-05	792	AQ779894	HS_3104_A1_E05_77C_CIT		gb_est85:BF243230	160.50	238.45	0.0004	1736	BF243230	601877381F1	NIH_MGC
gb_est102:BG565715	161.00	248.94	9.5e-05	792	BG565715	601193728F1	NIH_MGC_77	gb_est97:BF115874	160.50	238.07	0.0004	1776	BF115874	602136865F1	NIH_MGC
gb_est72:BE265715	161.00	247.48	0.0001	891	BE265715	601193728F1	NIH_MGC_77	gb_est91:BF683695	160.50	238.03	0.0004	1782	BF683695	602139955F1	NIH_MGC
gb_gss16:AZ540454	161.00	246.83	0.0001	939	AZ540454	ENTDF66TF_Entamoeba_hn		gb_est97:BG115863	160.50	237.81	0.0004	1814	BG		



9b_est86:BF3413305	164.00	251.01	7.3e-05	1010	BF3413305	602013244F1	NCl_CGAP	9b_est102:BG5372228	163.00	247.71	0.0001	1149	BG5372228	602565257F1	NiH_MG
9b_est80:BE9510532	164.00	250.80	7.5e-05	1037	BE9510532	601501073F1	NiH_MGC_7	9b_gss16:AQ751902	163.00	247.48	0.0001	1171	AQ751902	HS_5568_B2_F09_T7A	
9b_est93:BF6961222	164.00	250.72	7.6e-05	1034	BF6961222	9630507C11.x2_C	rein	9b_gss16:AQ751902	163.00	247.00	0.0001	1217	AQ751902	602149272F2	NiH_MG
9b_gss18:AQ878940	164.00	249.90	8.4e-05	1104	AQ878940	HS_3153_A1_G12_T7C_CIT		9b_est82:BF9638732	163.00	246.75	0.0001	1242	BF9638732	601455438F1	NiH_MG
9b_est100:BG432435	164.00	247.84	0.0001	1304	BG432435	602495456F1	NiH_MGC_7	9b_est84:BF9638732	163.00	244.87	0.0002	1445	BF9638732	602269456F1	NiH_MG
9b_est86:BF316672	164.00	247.13	0.0001	1381	BF316672	601903239F1	NiH_MGC_7	9b_est84:BF2070660	163.00	244.50	0.0002	1489	BF2070660	601870146F1	NiH_MG
9b_est97:BG164035	164.00	246.98	0.0001	1398	BG164035	602341048F1	NiH_MGC_8	9b_est101:BG501044	163.00	244.24	0.0002	1520	BG501044	602546451F1	NiH_MG
9b_est96:BG164035	164.00	246.85	0.0001	1412	BG164035	602282516F1	NiH_MGC_8	9b_est96:BG106136	163.00	243.90	0.0002	1562	BG106136	602289962F1	NiH_MG
9b_est98:BG250581	164.00	246.45	0.0001	1458	BG250581	602362760F1	NiH_MGC_9	9b_est98:BG250581	163.00	243.82	0.0002	1588	BG250581	602380913F1	NiH_MG
9b_est79:BE783704	164.00	246.44	0.0001	1460	BE783704	601471101F1	NiH_MGC_9	9b_est84:BF164623	163.00	243.70	0.0002	1588	BF164623	602380913F1	NiH_MG
9b_est98:BG251709	164.00	245.80	0.0001	1537	BG251709	602363612F1	NiH_MGC_9	9b_est89:BF535907	163.00	243.64	0.0002	1595	BF535907	602051525F1	NiH_MG
9b_est98:BG288934	164.00	245.75	0.0001	1543	BG288934	602383884F1	NiH_MGC_9	9b_est98:BG292060	163.00	243.57	0.0002	1604	BG292060	602386342F1	NiH_MG
9b_est98:BF569033	164.00	245.46	0.0001	1580	BF569033	602184338F1	NiH_MGC_4	9b_est84:BF160561	163.00	243.34	0.0002	1635	BF160561	601768340F1	NiH_MG
9b_est98:BF120255	164.00	245.15	0.0002	1620	BF120255	601756179F1	NiH_MGC_4	9b_est89:BG292068	163.00	242.63	0.0002	1731	BG292068	602294834F1	NiH_MG
9b_est97:BG165667	164.00	245.03	0.0002	1635	BG165667	602341569F1	NiH_MGC_8	9b_est85:BF302409	163.00	242.44	0.0002	1757	BF302409	602031413F1	NiH_MG
9b_est84:BF137780	164.00	244.63	0.0002	1689	BF137780	601782425F1	NiH_MGC_9	9b_est89:BF122019	163.00	242.32	0.0002	1774	BF122019	602031413F1	NiH_MG
9b_est97:BG115876	164.00	244.57	0.0002	1697	BG115876	602316868F1	NiH_MGC_8	9b_est89:BF121837	163.00	242.15	0.0002	1799	BF121837	602350836F1	NiH_MG
9b_est84:BF179980	164.00	244.53	0.0002	1702	BF179980	601806325F1	NiH_MGC_9	9b_est97:BG115864	163.00	241.80	0.0002	1851	BG115864	602316847F1	NiH_MG
9b_est83:BF101169	164.00	244.47	0.0002	1711	BF101169	601754745F1	NiH_MGC_9	9b_est98:BG292068	163.00	241.74	0.0002	1860	BG292068	602386353F1	NiH_MG
9b_est84:BF144168	164.00	244.38	0.0002	1723	BF144168	601788866F1	NiH_MGC_9	9b_est91:BF694727	163.00	241.67	0.0002	1870	BF694727	602380806F2	NiH_MG
9b_est100:BG388280	164.00	244.15	0.0002	1756	BG388280	602413278F1	NiH_MGC_9	9b_est97:BG114698	163.00	241.61	0.0002	1879	BF114698	602315345F1	NiH_MG
9b_est97:BG121865	164.00	243.81	0.0002	1804	BG121865	602350878F1	NiH_MGC_9	9b_est86:BF345553	163.00	241.19	0.0003	1944	BF345553	602019134F1	NiH_MG
9b_est84:BF144191	164.00	243.75	0.0002	1813	BF144191	601778895F1	NiH_MGC_9	9b_est81:BF962643	163.00	240.92	0.0003	1986	BF962643	601656009F1	NiH_MG
9b_est84:BF166248	164.00	243.55	0.0002	1843	BF166248	601777009F1	NiH_MGC_9	9b_est89:BF535923	163.00	240.44	0.0003	2065	BF535923	602051543F1	NiH_MG
9b_est98:BG261851	164.00	243.41	0.0002	1864	BG261851	602376515F1	NiH_MGC_9	9b_est89:BF180635	163.00	240.25	0.0003	2097	BF180635	601808744F1	NiH_MG
9b_est98:BG259511	164.00	242.46	0.0002	2011	BG259511	602393295F1	NiH_MGC_9	9b_est98:BF180635	162.50	252.24	6.2e-05	745	BF180635	602493037F1	NiH_MG
9b_gss16:AQ781402	163.50	252.96	5.7e-05	806	AQ781402	HS_3117_A2_D04_T7C_CIT		9b_est100:BG426856	162.50	249.66	8.7e-05	917	BG426856	602493037F1	NiH_MG
9b_est41:AV760732	163.50	252.88	5.7e-05	811	AV760732	AV760732_MDS	Home sap	9b_gss16:AQ781402	162.50	249.11	9.2e-05	954	AQ781402	Tetraodon nigroviridis	
9b_est19:BF625278	163.50	252.58	6.0e-05	831	BF625278	HV_C6a0012B06f	Hordeum	9b_gss16:AQ781402	162.50	249.04	9.3e-05	959	AQ781402	HS_5506_A2_B02_SP6	
9b_est100:BG400881	163.50	252.29	6.2e-05	851	BG400881	602464916F1	NiH_MGC_7	9b_est79:BF795661	162.50	249.04	9.4e-05	964	BF795661	601590646F1	NiH_MG
9b_est100:BG426804	163.50	252.26	6.2e-05	853	BG426804	602492981F1	NiH_MGC_7	9b_est100:BG430680	162.50	248.27	0.0001	1026	BG430680	602498476F1	NiH_MG
9b_est101:BG495197	163.50	251.15	7.2e-05	933	BG495197	602542816F1	NiH_MGC_5	9b_est100:BG439948	162.50	248.03	0.0001	1046	BG439948	602441493F1	NiH_MG
9b_est95:BF8971095	163.50	250.95	8.2e-05	1019	BF8971095	602270661F1	NiH_MGC_8	9b_gss16:AQ743361	162.50	247.76	0.0001	1069	BF43361	HS_5387_B2_H05_SP6	
9b_gss18:AQ897168	163.50	249.89	8.4e-05	1032	AQ897168	HS_3144_A1_E09_T7C_CIT		9b_est97:BG113870	162.50	247.71	0.0001	1073	BF113870	60284322F1	NiH_MG
9b_est101:BG503530	163.50	249.73	8.6e-05	1046	BG503530	602560105F1	NiH_MGC_6	9b_est98:BG283513	162.50	247.38	0.0001	1102	BF283513	602407848F1	NiH_MG
9b_est81:BF964994	163.50	249.28	9.1e-05	1084	BF964994	601658771F1	NiH_MGC_6	9b_est79:BF791313	162.50	247.24	0.0001	1115	BF791313	601582881F1	NiH_MG
9b_est100:BG438242	163.50	249.02	9.4e-05	1107	BG438242	602490151F1	NiH_MGC_1	9b_est100:BG392163	162.50	246.87	0.0001	1149	BG392163	602410276F1	NiH_MG
9b_est101:BG476006	163.50	248.86	9.6e-05	1122	BG476006	602521142F1	NiH_MGC_2	9b_est81:BF963493	162.50	246.70	0.0001	1164	BF963493	601657276F1	NiH_MG
9b_est101:BG518801	163.50	248.55	0.00010	1150	BG518801	602578394F1	NiH_MGC_2	9b_est92:BF797722	162.50	246.38	0.0001	1195	BF797722	602257664F1	NiH_MG
9b_gss18:AQ898051	163.50	247.81	0.0001	1221	AQ898051	HS_3135_B1_B01_MR_CIT		9b_est83:BF099902	162.50	245.81	0.0001	1251	BF099902	601752054F1	NiH_MG
9b_est95:BF974324	163.50	247.58	0.0001	1244	BF974324	602243981F1	NiH_MGC_4	9b_est83:BF099902	162.50	245.71	0.0001	1261	BF099902	601752054F1	NiH_MG
9b_est86:BF111903	163.50	247.58	0.0001	1244	BF111903	602282727F1	NiH_MGC_8	9b_est97:BG121856	162.50	245.41	0.0001	1292	BF121856	602350864F1	NiH_MG
9b_est86:BF339931	163.50	246.84	0.0001	1320	BF339931	602036452F1	NiH_MGC_9	9b_est98:BG326851	162.50	245.04	0.0002	1331	BG326851	602425735F1	NiH_MG
9b_est102:BG575002	163.50	246.70	0.0001	1335	BG575002	602598103F1	NiH_MGC_8	9b_est98:BG255903	162.50	244.42	0.0002	1399	BG255903	602367547F1	NiH_MG
9b_est86:BF341528	163.50	246.09	0.0001	1402	BF341528	602013532F1	NiH_MGC_9	9b_est85:BF244620	162.50	243.80	0.0002	1471	BF244620	601862644F1	NiH_MG
9b_est84:BF159075	163.50	245.34	0.0002	1490	BF159075	601766392F1	NiH_MGC_9	9b_est101:BG475479	162.50	243.33	0.0002	1528	BF475479	602491474F1	NiH_MG
9b_est80:BF8882936	163.50	245.23	0.0002	1491	BF8882936	601505836F1	NiH_MGC_7	9b_est86:BF338638	162.50	243.13	0.0002	1552	BF338638	602034326F2	NiH_MG
9b_est86:BF340966	163.50	245.23	0.0002	1503	BF340966	602032816F1	NiH_MGC_9	9b_est84:BF342987	162.50	242.99	0.0002	1570	BF342987	602017093F1	NiH_MG
9b_est85:BF343572	163.50	245.12	0.0002	1516	BF343572	602014403F1	NiH_MGC_9	9b_est83:BF102733	162.50	242.93	0.0002	1578	BF102733	601646363F1	NiH_MG
9b_est85:BF214488	163.50	244.03	0.0002	1655	BF214488	601845731F1	NiH_MGC_5	9b_est84:BF137767	162.50	242.52	0.0002	1631	BF137767	601782411F1	NiH_MG
9b_est84:BF144118	163.50	243.52	0.0002	1725	BF144118	601788808F1	NiH_MGC_5	9b_est83:BF133637	162.50	242.40	0.0002	1647	BF133637	601778529F1	NiH_MG
9b_est86:BF344985	163.50	243.33	0.0002	1752	BF344985	602014318F1	NiH_MGC_9	9b_est95:BF983387	162.50	242.32	0.0002	1657	BF983387	602305280F1	NiH_MG
9b_est89:BF579111	163.50	243.28	0.0002	1758	BF579111	602096045F1	NiH_MGC_9	9b_est83:BF131605	162.50	242.31	0.0002	1658	BF131605	601820411F1	NiH_MG
9b_est91:BF686973	163.50	243.02	0.0002	1796	BF686973	602102748F1	NiH_MGC_9	9b_est85:BF241521	162.50	242.28	0.0002	1662	BF241521	601878405F1	NiH_MG
9b_est82:BF035185	163.50	242.95	0.0002	1806	BF035185	601457042F1	NiH_MGC_9	9b_est98:BG262039	162.50	242.07	0.0002	1691	BG262039	602373873F1	NiH_MG
9b_est84:BF206094	163.50	242.80	0.0002	1828	BF206094	601869444F1	NiH_MGC_1	9b_est98:BG262039	162.50	241.95	0.0002	1708	BG262039	602373873F1	NiH_MG
9b_est97:BG165666	163.50	242.65	0.0002	1850	BF165666	602345168F1	NiH_MGC_8	9b_est83:BF099586	162.50	241.90	0.0002	1714	BF099586	601751605F1	NiH_MG
9b_est83:BF120228	163.50	241.89	0.0002	1967	BF120228	601756144F1	NiH_MGC_8	9b_est84:BF144112	162.50	241.74	0.0002	1737	BF144112	601788802F1	NiH_MG
9b_est83:BF134852	163.50	241.71	0.0002	1995	BF134852	601779236F1	NiH_MGC_9	9b_est84:BF164629	162.50	241.73	0.0002	1757	BF164629	601772410F1	NiH_MG
9b_est78:BF734690	163.50	241.42	0.0002	2043	BF734690	601570315F1	NiH_MGC_2	9b_est81:BF683686	162.50	241.60	0.0002	1758	BF683686	602139938F1	NiH_MG
9b_est97:BG115882	163.50	239.35	0.0003	2413	BF115882	602316879F1	NiH_MGC_8	9b_est84:BF137846	162.50	240.98	0.0003	1847	BF137846	601782456F1	NiH_MG
9b_gss17:AQ149085															



gb_est100:BG401417	167.00	256.92	3.4e-05	945	! BG401417 602464854F1 NIH_MGC_75	gb_gss30:AZ680982	165.00	255.04	4.3e-05	837	! A2680982 ENT1146TR Entamoeba
gb_est93:BF866440	167.00	256.05	3.8e-05	1011	! BF866440 963068H12.xl C. reinh	gb_est69:BE035369	165.00	254.36	4.7e-05	884	! BE035369 MM06H08 MM Mesembry
gb_est69:BG39530	167.00	256.05	3.8e-05	1014	! BG39530 AC06A12 AC Arabidopsi	gb_est101:GA925977	165.00	254.03	4.9e-05	908	! GA925977 60256359F1 NIH_MGC
gb_est85:BF248321	167.00	252.96	5.7e-05	1300	! BF248321 601821362F1 NIH_MGC_6	gb_est79:BE781618	165.00	253.66	5.2e-05	935	! BE781618 601467475F1 NIH_MGC
gb_est95:BG026360	167.00	252.88	5.7e-05	1309	! BG026360 60229130F1 NIH_MGC_8	gb_gss18:BF978460	165.00	253.44	5.3e-05	952	! BF978460 HS_3134_A2_G06_TTC
gb_est80:BE893481	167.00	252.45	6.1e-05	1355	! BE893481 601438029F1 NIH_MGC_7	gb_est95:BF976340	165.00	253.18	5.5e-05	972	! BF976340 602344471F1 NIH_MGC
gb_est91:BF6933957	167.00	251.41	6.9e-05	1473	! BF693395 602082536F1 NIH_MGC_8	gb_est97:BG117484	165.00	253.10	5.6e-05	978	! BG117484 602347606F1 NIH_MGC
gb_est85:BF240390	167.00	251.41	6.9e-05	1474	! BF240390 60109594F1 NIH_MGC_8	gb_est99:BG333926	165.00	252.28	6.2e-05	1045	! BG333926 60246264F1 NIH_MGC
gb_est86:BF345942	167.00	250.25	8.0e-05	1618	! BF345942 602017946F1 NIH_MGC_8	gb_est80:BE906076	165.00	252.07	6.4e-05	1063	! BE906076 601497110F1 NIH_MGC
gb_est79:BF788182	167.00	249.65	8.7e-05	1598	! BF788182 601482032F1 NIH_MGC_8	gb_est98:BG259857	165.00	250.04	8.2e-05	1252	! BG259857 602380004F1 NIH_MGC
gb_est98:BG288754	167.00	249.125	9.1e-05	1753	! BG288754 602385578F1 NIH_MGC_8	gb_est95:BF979734	165.00	249.97	8.3e-05	1259	! BF979734 602149244F2 NIH_MGC
gb_est95:BG032879	167.00	248.80	9.7e-05	1818	! BG032879 602300462F1 NIH_MGC_8	gb_est80:BE893389	165.00	249.64	8.7e-05	1293	! BE893389 60136719F1 NIH_MGC
gb_est91:BF676053	167.00	248.47	0.0001	1868	! BF676053 602084025F1 NIH_MGC_8	gb_est84:BF137783	165.00	249.08	9.3e-05	1353	! BF137783 601782428F1 NCICCG
gb_est98:BG297242	167.00	248.00	0.0001	1939	! BG297242 60234951F1 NIH_MGC_8	gb_est95:BF979425	165.00	248.24	0.0001	1447	! BF979425 602288255F1 NIH_MGC
gb_est92:BG027903	167.00	247.96	0.0001	1946	! BG027903 602294892F1 NIH_MGC_8	gb_est91:BF680542	165.00	248.06	0.0001	1468	! BF680542 602155280F1 NIH_MGC
gb_est82:BF033348	167.00	247.93	0.0001	1951	! BF033348 601438048F1 NIH_MGC_8	gb_est92:BF792678	165.00	246.79	0.0001	1627	! BF792678 602233862F1 NIH_MGC
gb_est43:AW155198	166.50	257.127	3.3e-05	858	! AW155198 mg1e0002B13f Rice bias	gb_est100:BG393495	165.00	246.77	0.0001	1636	! BG393495 602411875F1 NIH_MGC
gb_est40:BG434803	166.50	255.92	3.9e-05	957	! BG434803 602507324F1 NIH_MGC_79	gb_est100:BF979000	165.00	246.72	0.0001	1636	! BF979000 601588028F1 NIH_MGC
gb_gss18:BF79037	166.50	254.17	4.9e-05	1102	! BF79037 HS_3154_A2_H06_T7C C	gb_est89:BF572284	165.00	246.69	0.0001	1640	! BF572284 602077728F1 NIH_MGC
gb_est84:BF209842	166.50	253.63	5.2e-05	1151	! BF209842 60187537F1 NIH_MGC_5	gb_est80:BE888507	165.00	246.68	0.0001	1641	! BE888507 601512885F1 NIH_MGC
gb_gss3:CN502030	166.50	253.10	5.6e-05	1201	! AL206205 Tetraodon nigroviridi	gb_est95:BG034342	165.00	246.16	0.0001	1711	! BG034342 602302228F1 NIH_MGC
gb_est98:BG249535	166.50	252.86	5.7e-05	1224	! BG249535 602319592F1 NIH_MGC_8	gb_est98:BG256698	165.00	245.98	0.0001	1737	! BG256698 602307089F1 NIH_MGC
gb_est82:BF033674	166.50	252.77	5.8e-05	1233	! BF033674 602389213F1 NIH_MGC_5	gb_est89:BF532390	165.00	245.89	0.0001	1749	! BF532390 602074429F1 NCICCG
gb_est95:BG029415	166.50	252.61	5.9e-05	1249	! BG029415 602296847F1 NIH_MGC_8	gb_est101:BG438144	165.00	245.67	0.0001	1819	! BG438144 602052845F1 NIH_MGC
gb_est95:BG029415	166.50	251.91	6.5e-05	1322	! BG029415 602296847F1 NIH_MGC_8	gb_est86:BF338144	165.00	245.40	0.0001	1819	! BF338144 602037943F1 NCICCG
gb_est101:BG478460	166.50	251.81	6.5e-05	1329	! BG478460 602523912F1 NIH_MGC_8	gb_est84:BF180652	165.00	244.34	0.0002	1982	! BF180652 601808761F1 NCICCG
gb_est98:BG287280	166.50	251.84	6.6e-05	1332	! BG287280 602382218F1 NIH_MGC_9	gb_est80:BE909632	165.00	243.70	0.0002	2087	! BE909632 601495761F1 NIH_MGC
gb_est85:BF301197	166.50	251.34	7.0e-05	1384	! BF301197 602029615F1 NCICGAP	gb_gss21:AZ130339	164.50	256.66	3.5e-05	686	! AZ130339 OSNNB0104013f CG
gb_est83:BF120253	166.50	249.50	8.8e-05	1605	! BF120253 601756177F1 NCICGAP	gb_est80:BE875356	164.50	255.21	4.3e-05	771	! BE875356 601489068F1 NIH_MGC
gb_est80:BF2582186	166.50	249.37	9.0e-05	1622	! BE82186 601505287F2 NIH_MGC_7	gb_est102:AA849835	164.50	255.05	4.3e-05	781	! AA849835 ES1192602 Normal
gb_est85:BF214482	166.50	249.16	9.2e-05	1659	! BF214482 601485720F1 NIH_MGC_5	gb_gss3:CN502030	164.50	254.49	4.7e-05	817	! AL204580 Tetraodon nigrovir
gb_est80:BF724884	166.50	248.87	9.4e-05	1689	! BF724884 601447162F1 NIH_MGC_6	gb_est98:BG283225	164.50	254.28	4.8e-05	831	! BG283225 602406919F1 NIH_MGC
gb_est95:BG027865	166.50	248.77	9.6e-05	1689	! BG027865 602294837F1 NIH_MGC_8	em_estp110:BF631084	164.50	254.25	4.8e-05	833	! BF631084 HVSEB0014024f Hrd
gb_est84:BF180017	166.50	248.77	9.7e-05	1702	! BF180017 601806364F1 NCICGAP	gb_est101:BG533694	164.50	253.29	5.4e-05	900	! BG533694 602562348F1 NIH_MGC
gb_est94:BF965156	166.50	248.48	0.0001	1743	! BF965156 602286848F1 NIH_MGC_8	gb_est93:BF865294	164.50	252.85	5.7e-05	932	! BF865294 963058607.xl C. rei
gb_est21:AF1539812	166.00	265.08	1.2e-05	427	! AF1539812 tp57b01.xl NCICGAP	gb_gss16:AO7451439	164.50	252.26	6.2e-05	978	! AO7451439 HS_5573_B2_H02_Sp6
gb_est79:BE794234	166.00	258.55	2.8e-05	723	! BE794234 601591504F1 NIH_MGC_7	gb_gss16:AO7451439	164.50	251.98	6.4e-05	1000	! AO7451439 HS_5573_B2_H02_Sp6
gb_est100:BG426696	166.00	257.56	3.1e-05	783	! BG426696 602493241F1 NIH_MGC_75	gb_est96:BF810237	164.50	251.85	6.7e-05	1011	! BF810237 602282573F1 NIH_MGC
gb_gss3:CN502030	166.00	256.61	3.5e-05	845	! AL191091 Tetraodon nigroviridi	gb_est96:BF810237	164.50	251.65	6.7e-05	1027	! BF810237 602282573F1 NIH_MGC
gb_gss35:BO8887	166.00	256.18	3.8e-05	875	! BO8887 F207-Sp6 IGF Arabidopsis	gb_gss18:AO891386	164.50	251.43	6.9e-05	1045	! AO891386 HS_4831_B1_E08_7A
gb_est100:BE878848	166.00	254.46	4.7e-05	1005	! BE878848 601493141F1 NIH_MGC_8	gb_est89:BF537075	164.50	251.20	7.1e-05	1065	! BF537075 602049890F1 NCICCG
gb_est100:BG432101	166.00	253.94	5.0e-05	1048	! BG432101 602496965F1 NIH_MGC_8	gb_est101:BG436360	164.50	251.04	7.3e-05	1079	! BG436360 6020241861F1 NIH_MGC
gb_est100:BG429872	166.00	252.68	5.2e-05	1160	! BG429872 602494894F1 NIH_MGC_7	gb_est102:BG566496	164.50	250.72	7.6e-05	1107	! BG566496 602585519F1 NIH_MGC
gb_est81:BE966750	166.00	252.24	6.2e-05	1202	! BE966750 601661250R1 NIH_MGC_7	gb_est100:BG425865	164.50	250.70	7.6e-05	1109	! BG425865 602492225F1 NIH_MGC
gb_est98:BG294539	166.00	251.68	6.7e-05	1258	! BG294539 602391866F1 NIH_MGC_9	gb_est102:BG573646	164.50	249.55	8.8e-05	1216	! BG573646 602594668F1 NIH_MGC
gb_est97:BG114247	166.00	251.59	6.8e-05	1267	! BG114247 602285812F1 NIH_MGC_8	gb_est92:BF797946	164.50	248.78	9.7e-05	1294	! BF797946 602259866F1 NIH_MGC
gb_est99:BG334906	166.00	250.50	7.8e-05	1383	! BG334906 60241223F1 NIH_MGC_2	gb_est95:BF978745	164.50	248.74	9.7e-05	1298	! BF978745 602149267F2 NIH_MGC
gb_est91:BF683173	166.00	250.19	8.1e-05	1418	! BF683173 602139135F1 NIH_MGC_5	gb_est98:BG252799	164.50	248.43	0.0001	1331	! BG252799 602365367F1 NIH_MGC
gb_est86:BF248011	166.00	249.97	8.7e-05	1443	! BF248011 601859141F1 NIH_MGC_5	gb_est101:BG480437	164.50	248.23	0.0001	1353	! BG480437 602529659F1 NIH_MGC
gb_est89:BF338940	166.00	249.66	8.7e-05	1480	! BF338940 602051180F1 NCICGAP	gb_gss18:AO879040	164.50	247.52	0.0001	1433	! AO879040 HS_3154_A2_H09_7C
gb_est95:BG036535	166.00	249.53	8.8e-05	1496	! BF307176 601891309F1 NIH_MGC_1	gb_est98:BG295896	164.50	247.13	0.0001	1478	! BG295896 602393424F1 NIH_MGC
gb_est77:BE818335	166.00	249.39	9.0e-05	1512	! BG036535 602326303F1 NIH_MGC_9	gb_est86:BF342456	164.50	246.55	0.0001	1549	! BF342456 602013962F1 NCICCG
gb_est95:BG024018	166.00	247.94	0.0001	1698	! BG024018 601402325F1 NIH_MGC_6	gb_est93:BF859303	164.50	246.53	0.0001	1552	! BF859303 963001H01.yl C. rei
gb_est73:BF101072	165.50	246.40	0.0001	1924	! BF101072 601754636F1 NCICGAP	gb_est95:BG028429	164.50	246.34	0.0001	1575	! BG028429 602294333F1 NIH_MGC
gb_est100:BG304226	165.50	257.70	3.1e-05	723	! BG304226 601085045F1 NCICGAP	gb_est97:BF381058	164.50	246.27	0.0001	1584	! BF381058 601816246F1 NIH_MGC
gb_est100:BG403394	165.50	255.61	4.0e-05	856	! BG403394 602502107R1 NIH_MGC_75	gb_est93:BF859516	164.50	246.16	0.0001	1598	! BF859516 602155233F1 NIH_MGC
gb_est93:BF866469	165.50	255.46	4.1e-05	867	! BF866469 963069D03.xl C. reinh	gb_est86:BF344858	164.50	245.96	0.0001	1600	! BF344858 602014155F1 NCICCG
gb_est80:BE910542	165.50	254.75	4.5e-05	917	! BE910542 601501084F1 NIH_MGC_70	gb_est84:BF168005	164.50	245.88	0.0001	1625	! BF168005 602177455F1 NCICCG
gb_est98:BG0894095	165.50	254.27	4.7e-05	936	! BG0894095 HS_4832_A2_D10_T7C C	gb_est98:BG261841	164.50	245.38	0.0001	1702	! BG261841 602373640F1 NIH_MGC
gb_est102:BG0780609	165.50	254.27	4.8e-05	953	! BG0780609 HS_3138_A1_C10_T7C C	gb_est80:BE881292	164.50	245.22	0.0002	1724	! BE881292 601492375F1 NIH_MGC
gb_est99:BG339651	165.50	254.04	4.9e-05	971	! BG339651 60250312F1 NIH_MGC_4	gb_est95:BF232429	164.50	245.14	0.0002	1736	! BF232429 60242356F1 NIH_MGC
gb_est97:BG165672	165.50	251.41	6.9e-05	1200	! BG165672 602437219F1 NIH_MGC_4	gb_est85:BF217719	164.50	245.02	0.0002	1752	! BF217719 601882544F1 NIH_MGC
gb_est81:BE966575	165.50	249.52	8.8e-05	1398	! BE966575 602345176F1 NIH_MGC_8	gb_est84:BF206524	164.50	244.97	0.0002	1760	! BF206524 601869750F1 NIH_MGC
gb_est94:BF967885	165.50	248.56	0.00010	1510	! BF967885 601602518R1 NIH_MGC_7	gb_est98:BG214164	164.50	244.54	0.0002	1821	! BG214164 602315378F1 NIH_MGC
gb_est97:BG115852	165.50	247.41	0.0001	1657	! BF115852 602316826F1 NIH_MGC_8	gb_est85:BF3240750	164.50	243.71	0.0002	1947	! BF324075 601875340F1 NIH_MGC
gb_est85:BF6271253	165.50	247.17	0.0001	1689	! BF6271253 601844744F1 NIH_MGC_5	gb_est97:BG115879	164.50	243.59	0.0002	1966	! BF115879 602316873F1 NIH_MGC
gb_est93:BF866369	165.50	247.14	0.0001	1694	! BF866369 963016D07.xl C. reinh	gb_est97:BG114696	164.50	242.84	0.0002	2089	! BF114696 602315343F1 NIH_MGC
em_estp10:BF271174	165.50	246.09	0.0001	1843	! BF271174 GA_EB0010LH9f Gossyp	em_estp110:BF6293191	164.00	254.69	4.5e-05	751	! BF6293191 HVSEB00







[illegible]



91	est15: A1064123	+	334.50	545.14	3.0e-21	675	!	A1064123	GH04388.5prime	GH Dros	gb_est93: BF859150	+	190.00	291.56	4.0e-07	1343	!	BF859150	963001C06.x4	C. re
91	est18: BE0706461	-	339.00	556.57	7.0e-22	219	!	BE0706461	RC1-HT0268-170700-019		gb_est93: BF125422	+	190.00	289.43	5.3e-07	1595	!	BF125422	601763548F1	NIH_MG
91	est78: BF597655	+	323.00	547.19	2.3e-21	270	!	BF597655	R1KEN full-1e		gb_est93: BF165623	+	190.00	288.36	6.1e-07	1739	!	BF165623	602345106F1	NIH_MG
91	est96: BG070903	+	327.00	540.37	5.6e-21	358	!	BG070903	32-NIA Mouse		gb_est93: BF101057	+	189.50	286.11	7.9e-07	1916	!	BF101057	601754621F1	NCI_CG
91	est70: BE090826	+	324.00	540.20	5.6e-21	238	!	BE090826	PM2-BT0723-200300-001		gb_est93: BF180633	+	189.50	286.18	8.0e-07	1936	!	BF180633	601808742F1	NCI_CG
91	est30: A00117250	+	322.00	532.17	1.6e-20	348	!	A00117250	A00117250 Mouse two-cell		gb_est100: BG420926	+	189.00	292.92	3.4e-07	1050	!	BG420926	6024511337F1	NIH_MG
91	est55: B0031702	+	321.00	532.52	1.5e-20	295	!	B0031702	B0031702 RIKEN full-1e		gb_est100: BF570080	+	188.50	292.42	1.0e-06	1945	!	BF570080	602185905F1	NIH_MG
91	est41: A0749063	+	321.00	530.05	2.1e-20	360	!	A0749063	AV749063 NPC Homo sapi		gb_est86: BF345572	+	188.50	284.13	1.0e-06	1992	!	BF345572	602019164F1	NCI_CG
91	est86: BF364559	-	317.00	519.04	8.6e-20	506	!	BF364559	PM2-NN1083-280500-003		gb_est97: BF165643	+	187.50	284.96	9.4e-07	1625	!	BF165643	602345138F1	NIH_MG
91	est96: BG083741	+	315.00	519.30	8.3e-20	377	!	BG083741	H3091E11-5 NIA Mouse		gb_est93: BF124754	+	187.00	285.02	9.3e-07	1510	!	BF124754	601761441F1	NCI_CG
91	est16: A1135147	+	308.00	500.76	8.9e-19	691	!	A1135147	GH12788.5prime GH Dros		gb_est85: BF240341	+	186.50	288.53	5.9e-07	1063	!	BF240341	6020361026F1	NCI_CG
91	est95: BF930164	+	303.00	504.30	5.7e-19	245	!	BF930164	CM3-CN0092-251000-410		gb_est98: BG249311	+	186.00	285.30	9.0e-07	1288	!	BG249311	602361575F1	NIH_MG
91	est10: A0687846	+	303.00	497.58	1.3e-18	421	!	A0687846	nv09a11.s1 NCI_CGAP		gb_est100: BG436135	+	186.00	283.75	1.1e-06	1459	!	BG436135	602508536F1	NIH_MG
91	est31: A00178503	+	297.00	486.65	1.5e-18	200	!	A00178503	A00178503 Medaka ovary		gb_est80: BE9006514	+	186.00	281.98	1.4e-06	1683	!	BE9006514	601498379F1	NIH_MG
91	est10: A0681349	+	291.00	479.75	1.3e-17	368	!	A0681349	vr77F08.s1 Knowles Sol		gb_est84: BF1606314	+	186.00	279.90	1.8e-06	1990	!	BF1606314	601777089F1	NCI_CG
91	est19: A1763317	+	288.00	478.18	1.6e-17	259	!	A1763317	EST000039 HeLa SRG (S		gb_est98: BG283942	+	185.50	280.58	9.2e-07	1227	!	BG283942	602408112F1	NIH_MG
91	est55: B0032371	+	286.00	473.91	2.8e-17	278	!	B0032371	B0032371 RIKEN full-1e		gb_est84: BF138664	+	185.50	280.15	1.6e-06	1760	!	BF138664	601781845F1	NCI_CG
91	est64: B0423438	+	282.00	464.23	9.7e-17	351	!	B0423438	BB423438 RIKEN full-1e		gb_est85: BF244584	+	185.50	280.18	1.7e-06	1817	!	BF244584	601862601F1	NIH_MG
91	est41: A0006045	-	280.00	464.23	9.7e-17	351	!	A0006045	wz81B09.x1 NCI_CGAP		gb_est89: BF569058	+	185.50	279.54	1.9e-06	1914	!	BF569058	602186424F1	NCI_CG
91	est68: B05809397	+	282.00	456.33	6.5e-17	208	!	B05809397	BF589397 RIKEN full-1e		gb_est84: BF179987	+	185.50	279.70	2.1e-06	2047	!	BF179987	601806342F1	NIH_MG
91	est104: F12354	+	276.00	455.36	3.0e-16	316	!	F12354	HSC38H11 normalized inh		gb_gss20: AZ020428	+	185.00	294.01	2.9e-07	557	!	AZ020428	RPCI-2-335H4	TV RP
91	est21: A1532316	+	274.00	445.89	1.0e-15	552	!	A1532316	SD037719.5prime SD Dros		gb_est102: BG546853	+	185.00	292.92	3.4e-07	606	!	BG546853	602574035F1	NIH_MG
91	est82: BF007848	+	274.00	453.42	3.9e-16	281	!	BF007848	1589716 Amblyomma amer		gb_est97: BG115853	+	185.00	279.18	2.0e-06	1840	!	BG115853	602316828F1	NIH_MG
91	est84: BF149351	+	274.00	450.25	5.8e-16	363	!	BF149351	28.6 Human Epidermal k		gb_est95: BG035008	+	184.50	281.98	1.4e-06	1371	!	BG035008	602325282F1	NIH_MG
91	est54: A0931219	+	274.00	444.09	1.3e-15	596	!	A0931219	RDEIV28 Human rhabdomy		gb_est89: BF525803	+	184.50	280.60	1.6e-06	1532	!	BF525803	602069950F1	NCI_CG
91	est106: B083765	+	268.00	446.80	9.0e-16	211	!	B083765	KK2745F Human fetal hear		gb_est84: BF120646	+	184.00	279.97	1.8e-06	1506	!	BF120646	601758422F1	NCI_CG
91	est83: BF085557	-	262.00	435.24	4.0e-15	236	!	BF085557	MR2-GN0030-190900-008		gb_est97: BF165641	+	184.00	279.97	2.0e-06	1600	!	BF165641	602345136F1	NIH_MG
91	est54: A0936752	+	259.00	420.98	2.5e-14	529	!	A0936752	QV3-BN0047-230200-100		gb_est102: BG575872	+	184.00	278.86	2.0e-06	1647	!	BG575872	602598690F1	NIH_MG
91	est2: A0074000	+	259.00	419.14	3.1e-14	573	!	A0074000	mm82408.1 RIKEN full-1e		gb_est98: BG293855	+	183.50	280.27	1.7e-06	1373	!	BG293855	602393036F1	NIH_MG
91	est68: B0585916	-	256.00	423.81	1.7e-14	261	!	B0585916	BB585916 RIKEN full-1e		gb_est100: BG426850	+	183.00	284.72	9.7e-07	896	!	BG426850	602493031F1	NIH_MG
91	est84: A02861628	-	253.00	405.25	1.9e-13	773	!	A02861628	ZM0168C03F Mouse 10kb		gb_est100: BG404831	+	183.00	276.83	2.3e-06	1525	!	BG404831	6024210154F1	NIH_MG
91	est55: C0505848	-	248.50	396.53	5.7e-13	844	!	C0505848	26Tetraodon nigroviridis		gb_est100: BF861219	+	183.00	278.12	2.7e-06	1692	!	BF861219	963021A10.x1	C. re
91	est71: B0050448	+	246.00	391.63	1.1e-12	890	!	B0050448	Tetraodon nigroviridis		gb_gss12: AG529282	+	182.50	289.81	5.0e-07	555	!	AG529282	RPCI-11-367B23	TV R
91	est68: B0586125	+	245.00	404.79	2.0e-13	269	!	B0586125	BF586125 RIKEN full-1e		gb_est100: BG425897	+	182.50	282.14	1.3e-06	1030	!	BG425897	602492494F1	NIH_MG
91	est2: A02090343	-	243.00	394.27	7.6e-13	477	!	A02090343	RPCI-23-445C4.TJ RPCI		gb_est98: BG296116	+	182.00	271.73	3.9e-07	1938	!	BG296116	602393030F1	NIH_MG
91	est4: C0504488	+	242.00	384.77	2.6e-12	896	!	C0504488	Tetraodon nigroviridis		gb_est52: AW853357	+	182.00	291.29	3.7e-07	444	!	AW853357	RC1-CT0252-300999-0	
91	est102: BG598664	+	239.00	382.43	3.5e-12	718	!	BG598664	EST053564 CSTS Soltanum		gb_gss16: A0743307	+	182.00	280.69	1.6e-06	1081	!	A0743307	HS-5387	B2_B04
91	est106: N58565	+	235.00	377.71	2.9e-12	373	!	N58565	yv55B09.r1 Soares fetal		gb_est98: BG259344	+	182.00	278.49	2.1e-06	1291	!	BG259344	602378422F1	NIH_MG
91	est33: B0337559	+	235.00	377.71	6.4e-12	608	!	B0337559	ZM0132G24R Mouse 10kb		gb_est98: BF709592	+	182.00	277.88	2.3e-06	1356	!	BF709592	602076155F1	NIH_MG
91	est71: B0050448	+	229.00	374.85	9.2e-12	361	!	B0050448	RC1-HT0268-280200-015		gb_est98: BG436376	+	182.00	275.66	3.1e-06	1621	!	BG436376	601444337F1	NIH_MG
91	est16: A0779594	+	228.00	364.16	3.6e-11	696	!	A0779594	HS-5567_B2_D03.SP6 RPO		gb_est92: BF682951	+	182.00	274.43	3.6e-06	1791	!	BF682951	602109036F1	NCI_CG
91	est99: BG364469	-	225.00	367.37	2.4e-11	382	!	BG364469	dab95d08.y1 NICHD XGC		gb_est97: BG1218759	+	182.00	273.37	4.1e-06	1950	!	BG1218759	602350838F1	NIH_MG
91	est20: A0203699	-	225.00	358.03	8.0e-11	757	!	A0203699	RPCI-23-285A19.TJ RPCI		gb_est83: BF118759	+	182.00	273.87	4.4e-06	2030	!	BF118759	601755169F1	NCI_CG
91	est22: A1604717	+	223.00	363.46	4.0e-11	372	!	A1604717	vd71a12.y1 Beddington		gb_est84: BF179978	+	182.00	272.41	4.7e-06	2107	!	BF179978	601806323F1	NCI_CG
91	est7: A00399798	+	223.00	360.77	5.6e-11	462	!	A00399798	vd71a12.y1 Beddington		gb_est97: BG115847	+	181.50	273.88	3.9e-06	1748	!	BG115847	602316814F1	NIH_MG
91	est77: B0634375	+	223.00	358.38	7.6e-11	560	!	B0634375	uv67a05.y1 Soares mous		gb_est97: BG165648	+	181.50	273.88	3.9e-06	1748	!	BG165648	602345144F1	NIH_MG
91	est44: A0242979	-	221.00	355.51	1.1e-10	537	!	A0242979	gm91g07.x1 NCI_CGAP_k		gb_est84: BF166258	+	181.50	273.20	4.2e-06	1086	!	BF166258	602177021F1	NCI_CG
91	est29: A0586843	+	220.00	353.38	1.4e-10	556	!	A0586843	AL586843 Stratagene CH		gb_est100: BG400929	+	181.00	278.97	2.0e-06	1843	!	BG400929	602464968F1	NIH_MG
91	est79: BF780744	+	212.50	326.69	4.4e-09	1715	!	BF780744	601469616F1 NIH_MGC		gb_est94: BF964548	+	181.00	277.01	2.3e-06	1170	!	BF964548	602267844F1	NIH_MG
91	est106: N58572	+	212.00	346.25	3.6e-10	331	!	N58572	yv55D09.r1 Soares fetal		gb_est101: BG533805	+	181.00	277.08	2.6e-06	1261	!	BG533805	602562478F1	NIH_MG
91	est21: A1530941	+	209.50	336.65	1.2e-09	510	!	A1530941	SD017592.5prime SD Dros		gb_est101: BF8660716	+	181.00	274.81	3.4e-06	1514	!	BF8660716	963017D07.x1	C. re
91	est21: A1534657	+	209.50	336.62	1.2e-09	511	!	A1534657	SD07592.5prime SD Dros		gb_est99: BG342563	+	181.00	274.60	3.5e-06	1541	!	BG342563	602374740F1	NIH_MG
91	est84: BF206496	+	202.50	309.15	4.2e-08	1797	!	BF206496	601869704F1 NIH_MGC		gb_est89: BF572295	+	181.00	273.51	4.1e-06	1682	!	BF572295	602077437F1	NIH_MG
91	est11: A0749512	+	202.00	338.01	1.0e-09	164	!	A0749512	HGBB7047 Human Gliabla		gb_est84: BF158995	+	181.00	272.80	4.5e-06	1781	!	BF158995	601766305F1	NCI_CG
91	est4: C0504589	+	202.00	324.99	5.5e-09	468	!	C0504589	AT275166 Tetraodon nigroviridis		gb_est97: BG114717	+	181.00	271.97	5.0e-06	1904	!	BG114717	602315387F1	NIH_MG
91	est85: BF214484	+	201.00	305.82	6.4e-08	1914	!	BF214484												



OM of: US-09-512-581-2 to: EST: \* out\_format : pfs

Date: Sep 25, 2001 11:28 PM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

MODEL=framet.p2n.model -DEV=slp  
Q/cgn2-1/USPto\_Spool/US09512581/runat\_25092001\_095746\_19380/app\_query.fasta\_1.1470  
DB=EST -QWMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
MINMATCH=0.000 -LOEPC=0.000 -LOPEXT=0.000 -QGAPOP=4.500  
QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
DELEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi  
LIST=1000 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=50 -MODB=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09512581\_@CGN1\_1.4971 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-512-581-2

Query length: 1391

Database: EST \*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 2155.920000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est88:BF509252	+ 1225.00	2053.29	3.0e-105	791	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509253	+ 1143.50	1916.01	1.3e-97	734	! BF509253 602370939F1 NIH_MGC_92
gb_est88:BF509254	+ 1091.00	1827.07	1.2e-92	729	! BF509254 602370939F1 NIH_MGC_75
gb_est88:BF509255	+ 1066.00	1781.47	4.1e-90	944	! BF509255 601788573F1 NIH_MGC_91
gb_est88:BF509256	+ 1061.50	1775.17	9.1e-90	848	! BF509256 601788573F1 NIH_MGC_69
gb_est88:BF509257	+ 1027.00	1721.04	9.6e-87	596	! BF509257 602142496F1 NIH_MGC_92
gb_est88:BF509258	+ 1023.00	1710.57	3.7e-86	802	! BF509258 602142496F1 NIH_MGC_92
gb_est88:BF509259	+ 974.50	1630.32	1.1e-81	683	! BF509259 60123013F1 NIH_MGC_92
gb_est88:BF509260	+ 969.00	1622.43	3.0e-81	608	! BF509260 602142496F1 NIH_MGC_92
gb_est88:BF509261	+ 953.00	1593.50	1.2e-79	703	! BF509261 602142496F1 NIH_MGC_92
gb_est88:BF509262	+ 949.00	1589.14	2.1e-79	578	! BF509262 602142496F1 NIH_MGC_92
gb_est88:BF509263	+ 944.00	1577.68	9.1e-79	735	! BF509263 602142496F1 NIH_MGC_92
gb_est88:BF509264	+ 932.00	1555.06	1.7e-77	883	! BF509264 602142496F1 NIH_MGC_92
gb_est88:BF509265	+ 927.00	1539.97	1.2e-76	1504	! BF509265 602142496F1 NIH_MGC_92
gb_est88:BF509266	+ 919.00	1535.88	2.0e-76	701	! BF509266 602142496F1 NIH_MGC_92
gb_est88:BF509267	+ 899.00	1498.83	2.3e-74	902	! BF509267 602142496F1 NIH_MGC_92
gb_est88:BF509268	+ 893.00	1491.15	6.1e-74	738	! BF509268 602142496F1 NIH_MGC_92
gb_est88:BF509269	+ 890.00	1488.46	8.6e-74	608	! BF509269 602142496F1 NIH_MGC_92
gb_est88:BF509270	+ 884.00	1473.63	5.8e-73	885	! BF509270 602142496F1 NIH_MGC_92
gb_est88:BF509271	+ 850.00	1422.97	3.8e-70	662	! BF509271 602142496F1 NIH_MGC_92
gb_est88:BF509272	+ 850.00	1415.96	9.4e-70	886	! BF509272 602142496F1 NIH_MGC_92
gb_est88:BF509273	+ 830.00	1387.73	3.5e-68	560	! BF509273 602142496F1 NIH_MGC_92
gb_est88:BF509274	+ 828.00	1386.26	4.3e-68	480	! BF509274 602142496F1 NIH_MGC_92
gb_est88:BF509275	+ 827.00	1377.47	1.3e-67	910	! BF509275 602142496F1 NIH_MGC_92
gb_est88:BF509276	+ 817.00	1366.26	5.5e-67	535	! BF509276 602142496F1 NIH_MGC_92
gb_est88:BF509277	+ 816.00	1365.45	6.0e-67	490	! BF509277 602142496F1 NIH_MGC_92
gb_est88:BF509278	+ 806.00	1343.48	1.0e-65	746	! BF509278 602142496F1 NIH_MGC_92
gb_est88:BF509279	+ 806.00	1341.31	1.4e-65	888	! BF509279 602142496F1 NIH_MGC_92
gb_est88:BF509280	+ 797.50	1330.81	5.2e-65	648	! BF509280 602142496F1 NIH_MGC_92
gb_est88:BF509281	+ 786.00	1315.29	3.8e-64	470	! BF509281 602142496F1 NIH_MGC_92
gb_est88:BF509282	+ 783.00	1307.60	1.0e-63	580	! BF509282 602142496F1 NIH_MGC_92
gb_est88:BF509283	+ 770.00	1285.77	1.7e-62	570	! BF509283 602142496F1 NIH_MGC_92
gb_est88:BF509284	+ 767.00	1275.07	6.6e-62	896	! BF509284 602142496F1 NIH_MGC_92
gb_est88:BF509285	+ 751.00	1247.29	2.3e-60	944	! BF509285 602142496F1 NIH_MGC_92
gb_est88:BF509286	+ 747.00	1242.92	4.1e-60	777	! BF509286 602142496F1 NIH_MGC_92
gb_est88:BF509287	+ 746.00	1248.45	2.0e-60	434	! BF509287 602142496F1 NIH_MGC_92
gb_est88:BF509288	+ 741.00	1239.89	6.0e-60	437	! BF509288 602142496F1 NIH_MGC_92
gb_est88:BF509289	+ 739.50	1233.51	1.4e-59	595	! BF509289 602142496F1 NIH_MGC_92
gb_est88:BF509290	+ 731.00	1223.33	5.1e-59	423	! BF509290 602142496F1 NIH_MGC_92
gb_est88:BF509291	+ 731.00	1222.70	5.5e-59	445	! BF509291 602142496F1 NIH_MGC_92
gb_est88:BF509292	+ 726.00	1211.68	2.3e-58	546	! BF509292 602142496F1 NIH_MGC_92

gb_est13:AA913927	717.00	1197.94	1.3e-57	483	! AA913927 0139f02.sl Soares_N
gb_est13:AA913928	717.00	1197.87	1.3e-57	486	! AA913928 0139f02.sl Soares_N
gb_est13:AA913929	680.50	1133.78	4.9e-54	580	! BF451878 601824203F1 NIH_MGC
gb_est13:AA913930	678.50	1126.12	1.3e-53	818	! BF107038 601824203F1 NIH_MGC
gb_est13:AA913931	677.50	1125.48	1.4e-53	751	! AA920204 RVC1-BN0047-300800-2
gb_est13:AA913932	677.00	1128.98	9.1e-54	529	! BF292587 QV3-BN0047-300800-2
gb_est13:AA913933	673.00	1122.95	2.0e-53	498	! BF558326 678g06.sl Zebrafish
gb_est13:AA913934	670.00	1116.47	4.5e-53	557	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913935	667.00	1115.78	4.9e-53	391	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913936	667.00	1106.99	1.5e-52	794	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913937	658.00	1098.61	4.5e-52	456	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913938	649.00	1084.72	2.7e-51	408	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913939	645.00	1070.03	1.7e-50	772	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913940	639.00	1067.14	2.5e-50	429	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913941	638.00	1063.55	4.0e-50	500	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913942	634.00	1049.86	2.3e-49	872	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913943	629.00	1050.96	2.0e-49	403	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913944	616.50	1026.99	4.4e-48	504	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913945	616.00	1026.37	3.7e-48	421	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913946	577.50	955.02	4.5e-44	806	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913947	576.00	958.94	2.7e-44	479	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913948	574.50	958.14	3.0e-44	416	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913949	567.50	942.30	2.3e-43	573	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913950	557.50	920.25	3.9e-42	884	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913951	553.50	913.11	9.6e-42	889	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913952	549.50	904.57	2.9e-41	1024	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913953	548.00	906.09	2.4e-41	738	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913954	545.50	902.28	3.9e-41	713	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913955	534.00	884.44	3.8e-40	414	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913956	521.00	870.86	2.2e-39	475	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913957	514.50	848.87	3.6e-38	763	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913958	513.50	850.28	3.0e-38	594	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913959	512.50	843.04	7.7e-38	929	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913960	507.00	846.82	4.7e-38	333	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913961	503.00	837.89	1.5e-37	384	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913962	498.00	829.64	4.3e-37	377	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913963	496.50	816.43	2.3e-36	890	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913964	494.00	822.82	1.0e-36	378	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913965	469.00	779.91	2.5e-34	394	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913966	458.00	765.90	1.5e-33	271	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913967	456.00	756.24	5.3e-33	449	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913968	455.00	759.07	3.7e-33	312	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913969	454.00	757.25	4.6e-33	335	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913970	452.00	752.98	8.0e-33	338	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913971	448.00	741.85	3.3e-32	480	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913972	442.00	737.51	1.8e-32	300	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913973	430.50	707.60	2.7e-30	543	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913974	428.50	704.57	4.0e-30	674	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913975	428.00	708.25	2.5e-30	468	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913976	419.50	685.23	4.7e-29	936	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913977	416.00	692.01	2.0e-29	336	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913978	416.00	681.69	7.5e-29	742	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913979	410.00	674.61	1.9e-28	674	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913980	410.00	681.80	7.4e-29	337	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913981	407.00	677.16	1.3e-28	325	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913982	395.50	643.70	9.8e-27	1001	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913983	393.00	633.69	2.7e-27	318	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913984	391.00	639.08	1.8e-26	785	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913985	383.50	628.82	6.2e-26	644	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913986	380.50	630.02	5.6e-26	388	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913987	375.00	618.48	2.5e-25	464	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913988	374.00	623.78	1.3e-25	264	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913989	374.00	617.87	2.7e-25	425	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913990	373.50	604.79	1.4e-24	1139	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913991	372.00	618.23	2.3e-25	296	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913992	372.00	618.23	2.6e-25	314	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913993	370.00	603.47	1.7e-24	785	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913994	353.00	585.47	1.7e-23	328	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913995	341.00	557.10	6.5e-22	626	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913996	340.00	555.41	8.1e-22	626	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913997	337.00	545.37	2.9e-21	933	! BF306904 601807447F1 NIH_MGC
gb_est13:AA913998	334.50	546.04	2.7e-21	628	! BF306904 601807447F1 NIH_MGC



**THIS PAGE BLANK (USPTO)**







## JOURNAL

Submitted (23-FEB-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)

## COMMENT

On May 31, 2000 this sequence version replaced gi:7106128.

----- Genome Center  
Center: RIKEN Genomic Sciences Center(GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@gs.c.riken.go.jp  
----- Project Information  
Center project name: HumDraft11  
Center clone name: CMB9-97D23  
----- Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 117712 bases at least Q40  
Consensus quality: 123262 bases at least Q30  
Consensus quality: 125535 bases at least Q20  
Insert size: 127022; sum-of-contigs  
Quality coverage: 5.18x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 18938 contig of 18938 bp in length
19039 34051 contig of 15013 bp in length
34152 45991 contig of 11840 bp in length
46092 58090 contig of 11999 bp in length
58191 68410 contig of 10220 bp in length
68511 76996 contig of 8486 bp in length
77097 86120 contig of 9024 bp in length
86221 92736 contig of 6516 bp in length
92837 98615 contig of 5779 bp in length
98716 103151 contig of 4436 bp in length
103252 107361 contig of 4110 bp in length
107462 111521 contig of 4060 bp in length
111622 115356 contig of 3735 bp in length
115457 118625 contig of 3169 bp in length
118726 121644 contig of 2919 bp in length
121745 123829 contig of 2085 bp in length
123930 125530 contig of 1601 bp in length
125631 127610 contig of 1980 bp in length
127711 128822 contig of 1112 bp in length

```

Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 18938 contig of 18938 bp in length
18939 19038 gap of 100 bp
19039 34051 contig of 15013 bp in length
34052 34151 gap of 100 bp
34152 45991 contig of 11840 bp in length
45992 46091 gap of 100 bp
46092 58090 contig of 11999 bp in length
58091 58190 gap of 100 bp
58191 68410 contig of 10220 bp in length
68411 68510 gap of 100 bp
68511 76996 contig of 8486 bp in length
76997 77096 gap of 100 bp
77097 86120 contig of 9024 bp in length

```

```

* 86121 86220 gap of 100 bp
86221 92736 contig of 6516 bp in length
92737 92836 gap of 100 bp
92837 98615 contig of 5779 bp in length
98616 98715 gap of 100 bp
98716 103151 contig of 4436 bp in length
103152 103251 gap of 100 bp
103252 107361 contig of 4110 bp in length
107362 107461 gap of 100 bp
107462 111521 contig of 4060 bp in length
111522 111621 gap of 100 bp
111622 115356 contig of 3735 bp in length
115357 115456 gap of 100 bp
115457 118625 contig of 3169 bp in length
118626 118725 gap of 100 bp
118726 121644 contig of 2919 bp in length
121645 121744 gap of 100 bp
121745 123829 contig of 2085 bp in length
123830 123929 gap of 100 bp
123930 125530 contig of 1601 bp in length
125531 125630 gap of 100 bp
125631 127610 contig of 1980 bp in length
127611 127710 gap of 100 bp
127711 128822 contig of 1112 bp in length.

```

## FEATURES

## Source

```

1. .128822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="liql2"
/clone="CMB9-97D23"

```

```

misc_feature 1. .18938
/note="assembly_fragment"
misc_feature 19039. .34051
/note="assembly_fragment"
misc_feature 34152. .45991
/note="assembly_fragment"
misc_feature 46092. .58090
/note="assembly_fragment"
misc_feature 58191. .68410
/note="assembly_fragment"
misc_feature 68511. .76996
/note="assembly_fragment"
misc_feature 77097. .86120
/note="assembly_fragment"
misc_feature 86221. .92736
/note="assembly_fragment"
misc_feature 92837. .98615
/note="assembly_fragment"
misc_feature 98716. .103151
/note="assembly_fragment"
misc_feature 103252. .107361
/note="assembly_fragment"
misc_feature 107462. .111521
/note="assembly_fragment"
misc_feature 111622. .115356
/note="assembly_fragment"
misc_feature 115457. .118625
/note="assembly_fragment"
misc_feature 118726. .121644
/note="assembly_fragment"
misc_feature 121745. .123829
/note="assembly_fragment"
misc_feature 123930. .125530
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 125631. .127610
/note="assembly_fragment"
misc_feature 127711. .128822
/note="assembly_fragment"

```

BASE COUNT 36476 a 26110 c 27112 g 37324 t 1800 others  
ORIGIN











```

VERSION 295113.2 GI:6572188
KEYWORDS HTG; ca repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 110042)
JOURNAL Direct Submission
COMMENT Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:2253036.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em; EMBL; Sw; SWISSPROT; Tr; TrEMBL; Wp; WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
CFA-175E3 is from the human BAC library described in U-J. Kim et
al. (1996) Genomics 34, 213-218.
VECTOR: pBAC108L
This sequence is the entire insert of clone CFA-175E3.
FEATURES
Source
Location/Qualifiers
1..110042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CFA-175E3"
/clone_lib="CIT978SK-A1"
28..515
/note="match: STS: Em:B14235; match: GSS: Em:B14235"
181..318
/note="FLAM_A repeat: matches 1..133 of consensus"
328..437
/note="L2 repeat: matches 2092..2206 of consensus"
584..868
/note="AluSg repeat: matches 1..280 of consensus"
1029..1644
/note="match: GSS: Em:B18037; match: STS: Em:B18037"
1194..1496
/note="AluSx repeat: matches 1..298 of consensus"
2176..2209
/note="17 copies 2 mer tt 88 conserved"
2931..3180
/note="AluJo repeat: matches 60..310 of consensus"
4076..4378
/note="AluJo repeat: matches 2..302 of consensus"
4413..4564
/note="MER91B repeat: matches 28..179 of consensus"
4885..5107
/note="AluSg/x repeat: matches 82..301 of consensus"
5182..5484
/note="AluSx repeat: matches 2..304 of consensus"
5632..5799
/note="AluSg/x repeat: matches 121..300 of consensus"
5875..6032
/note="MIR repeat: matches 21..195 of consensus"
8190..8485
/note="AluJo repeat: matches 1..299 of consensus"
complement(8360..8763)
/note="match: STS: Em:R08275"
complement(8421..8889)
/note="match: GSS: Em:AQ547651"
8638..9058
/note="match: GSS: Em:AQ219894"
8669..9019
/note="match: GSS: Em:AQ550076"
8895..9305
/note="match: GSS: Em:AQ347278"
9043..9156
/note="MIR repeat: matches 106..221 of consensus"
9292..9586
/note="Aluwb repeat: matches 1..298 of consensus"
9661..9872
/note="MIR repeat: matches 9..231 of consensus"
10043..10089
/note="L2 repeat: matches 2702..2749 of consensus"
10220..10541
/note="AluSg repeat: matches 1..310 of consensus"
12139..12445
/note="AluSg repeat: matches 1..306 of consensus"
12867..13141
/note="AluSx repeat: matches 1..263 of consensus"
13142..13238
/note="Aluwb repeat: matches 51..145 of consensus"
13288..13589
/note="AluSg repeat: matches 1..297 of consensus"
13913..14079
/note="L2 repeat: matches 2428..2602 of consensus"
14271..14346
/note="MIR repeat: matches 59..144 of consensus"
15560..15660
/note="L2 repeat: matches 1365..1458 of consensus"
15903..16213
/note="AluSg repeat: matches 2..311 of consensus"
17782..18082
/note="AluSg repeat: matches 1..302 of consensus"
18355..18736
/note="MLT1B repeat: matches 1..390 of consensus"
19322..19608
/note="AluJo repeat: matches 1..289 of consensus"
20518..20553
/note="18 copies 2 mer ac 100 conserved"
20822..20933
/note="FLAM_A repeat: matches 18..129 of consensus"
20824..20978
/note="BC200 repeat: matches 21..176 of consensus"
21715..21834
/note="6 copies 20 mer 79 conserved"
21720..21829
/note="55 copies 2 mer tg 84 conserved"
22568..22730
/note="MIR repeat: matches 87..256 of consensus"
22791..23055
/note="AluSx repeat: matches 1..298 of consensus"
23216..23514
/note="AluSg repeat: matches 1..299 of consensus"
23517..23819
/note="AluSx repeat: matches 1..303 of consensus"
24150..24208
/note="MIR repeat: matches 77..130 of consensus"
25160..25475
/note="MER46B repeat: matches 2..235 of consensus"
26167..26254
/note="MIR repeat: matches 61..151 of consensus"
26831..26980
/note="L1MC5 repeat: matches 7710..7869 of consensus"
27721..27992
/note="AluJo repeat: matches 5..289 of consensus"

```



BASE COUNT	31945 a	18533 c	18883 g	30840 t	724 others
ORIGIN					
Query Match	0.6%;	Score 30;	DB 83;	Length 100925;	
Best Local Similarity	100.0%;	Pred. No.	0.00023;		
Matches	30;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY	4428	gaaaaatctttttttttttttttttttttt 4457			
Dd	1637	GAAAAATCTTTTTTTTTTTTTTTTTTTT 1608			
RESULT 47					
AC073980/c					
LOCUS	AC073980	101539 bp	DNA	PRI	08-DEC-2000
DEFINITION	Homo sapiens clone RP11-312H11,	complete sequence.			
ACCESSION	AC073980				
VERSION	AC073980.5	GI:11597142			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 101539)				
JOURNAL	Waterston,R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 101539)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (08-JUL-2000) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
REFERENCE	3 (bases 1 to 101539)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-DEC-2000) Genome Sequencing Center, Washington				
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
AUTHORS	MO 63108, USA				
TITLE	4 (bases 1 to 101539)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (08-DEC-2000) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
COMMENT	On Dec 7, 2000 this sequence version replaced gi:11276284.				
FEATURES	Center project name: H_NH0312H11.				
source	Location/Qualifiers				
	1..101539				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="RP11-312H11"				
BASE COUNT	34020 a	18550 c	18384 g	30585 t	
ORIGIN					
Query Match	0.6%;	Score 30;	DB 88;	Length 101539;	
Best Local Similarity	100.0%;	Pred. No.	0.00023;		
Matches	30;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY	4428	gaaaaatctttttttttttttttttttttt 4457			
Dd	25994	GAAAAATCTTTTTTTTTTTTTTTTTTTT 25965			
RESULT 48					
HS175E3					
LOCUS	HS175E3	110042 bp	DNA	PRI	12-DEC-1999
DEFINITION	Human DNA sequence from clone CTA-175E3 on chromosome 22q12.1				
ACCESSION	Contains an exon of a novel gene, ESTs, STSS, CA repeat				
	polymorphism and GSSs, complete sequence.				
	295113				















FEATURES | source



```

RESULT 40
AC021197
LOCUS      AC021197 172105 bp  DNA               HTG              07-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-798B19, WORKING DRAFT
SEQUENCE 1, 21 unordered pieces.
ACCESSION AC021197.5 GI:8570404
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 172105)
            Waterston, R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 172105)
            Waterston, R.H.
            Direct Submission
            Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Jun 17, 2000 this sequence version replaced gi:7235351.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0798B19
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160434 bases at least Q40
Consensus quality: 164353 bases at least Q30
Consensus quality: 166763 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 170105; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1301: contig of 1301 bp in length
1302 1401: gap of unknown length
1402 4060: contig of 2659 bp in length
4061 4160: gap of unknown length
4161 7024: contig of 2864 bp in length
7025 7124: gap of unknown length
7125 9662: contig of 2538 bp in length
9663 9762: gap of unknown length
9763 13070: contig of 3308 bp in length
13071 13170: gap of unknown length
13171 15928: contig of 2758 bp in length
15929 16029: gap of unknown length
16029 19294: contig of 3266 bp in length
19295 19395: gap of unknown length
19395 23283: contig of 3889 bp in length
23284 23384: gap of unknown length
23384 27272: contig of 3889 bp in length
27273 27373: gap of unknown length
27373 32635: contig of 5163 bp in length
32635 40106: gap of unknown length
40106: contig of 7471 bp in length

```

```

FEATURES
            source
            1..172105
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="19"
            /clone="RP11-798B19"
            1..1301
            /note="assembly_name:Contig11"
            1402..4060
            /note="assembly_name:Contig12"
            4161..7024
            /note="assembly_name:Contig13"
            7125..9662
            /note="assembly_name:Contig14"
            9763..13070
            /note="assembly_name:Contig15"
            13171..15928
            /note="assembly_name:Contig16"
            16029..19294
            /note="assembly_name:Contig17"
            19395..23283
            /note="assembly_name:Contig18"
            23384..27272
            /note="assembly_name:Contig19"
            clone_end:SP6
            vector_side:right
            27373..32535
            /note="assembly_name:Contig20"
            32636..40106
            /note="assembly_name:Contig21"
            40207..45656
            /note="assembly_name:Contig22"
            45757..52232
            /note="assembly_name:Contig23"
            52333..59204
            /note="assembly_name:Contig24"
            59305..68409
            /note="assembly_name:Contig25"
            68510..76838
            /note="assembly_name:Contig26"
            76939..88902
            /note="assembly_name:Contig27"
            89003..100235
            /note="assembly_name:Contig28"
            100336..117760
            /note="assembly_name:Contig29"
            117861..142316
            /note="assembly_name:Contig30"
            142417..172105
            /note="assembly_name:Contig31"
            43255 a 42852 c 43254 g 40653 t 2091 others
BASE COUNT
ORIGIN

```











-2001



## JOURNAL

Submitted (22-NOV-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 156840)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Waterston,R.H.  
 Direct Submission

Submitted (05-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Waterston,R.  
 Direct Submission

Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On May 5, 1999 this sequence version replaced gi:4337251.

## COMMENT

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0215P16  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3 6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-21N8, 200 bp overlap. Actual start of this clone is at base position 111596 of CTB-21N8; actual end is at base position 156840 of RP11-215P16.

## FEATURES

## Source

1. 156840  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q11.2-q22"  
 /clone="RP11-215P16"  
 /clone.lib="RPCI-11"  
 552..938  
 /rpt\_family="L2"  
 2050..2156  
 /rpt\_family="Alu"

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## misc\_feature

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## misc\_feature

## gene

## CDS

2158..2427

/rpt\_family="Alu"

3750..3781

/rpt\_family="(CA)n"

3880..3991

/rpt\_family="L2"

3990..4122

/rpt\_family="MERL\_type"

4464..4601

/rpt\_family="Alu"

4602..4621

/rpt\_family="(CAA)n"

4622..4768

/rpt\_family="Alu"

9706..9767

/rpt\_family="AT-rich"

9796..10014

/rpt\_family="L1"

10032..10425

/note="match to EST AA780910 (NID:g2840241) ag98f04.sl"

10536..10573

/rpt\_family="AT-rich"

10758..11054

/rpt\_family="Alu"

11374..11465

/rpt\_family="L2"

12331..13075

/rpt\_family="L1"

13131..13252

/rpt\_family="Alu"

13405..13501

/note="similar to Mus musculus EST A1325828 (NID:g4060257)

mf66g02.yl"

<13409..>134473

/gene="WUGSC:H\_NH0215P16.1"

join(<13409..13501,49523..49585,67833..67956,

105929..106049,134380..>134473)

/gene="WUGSC:H\_NH0215P16.1"

/note="match to BAA74857.1 (PID:g4240157); H\_NH0215P16.1"

/codon\_start=1

/product="match to KIAA0834"

/protein\_id="AAF19245.1"

/db\_xref="GI:6624100"

/translation="SLKGLKHANIVLLHDIHTKETLTLVFEVHTDLCQYMDKHPC GLHPDNVKLEFLQRLGLSVLHORYILHRLDKPCNLISDTGELKLADFLGAKSVPSHYNSVEVTLWRPPDVLGSGTEYSTCLDWMGVGVCIFVEMIQGVAAFPGMKDIQDQL ERIRL"

13903..13945

/rpt\_family="(CA)n"

15936..16133

/rpt\_family="MIR"

16450..16506

/rpt\_family="(CA)n"

17822..17909

/rpt\_family="L1"

17935..18069

/rpt\_family="L1"

18076..18376

/rpt\_family="Alu"

18385..18490

/rpt\_family="L1"

18491..18570

/rpt\_family="(TA)n"

18575..18717

/rpt\_family="L1"

18911..19121

/rpt\_family="MIR"

19444..19954

/rpt\_family="L2"

20346..20736

/rpt\_family="L1"

21973..22023

/rpt\_family="MIR"











```

RESULT 33
AC022305
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-276K9 map 15q21, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC022305
VERSION AC022305.2 GI:8247803
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 94409)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 94409)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (29-JAN-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Jun 4, 2000 this sequence version replaced gi:6814954.
-----
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leetowen@systemsbiology.org
-----
Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
-----
* NOTE: This record contains 96 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
865: contig of 865 bp in length
866: gap of unknown length
966: contig of 845 bp in length
1811: gap of unknown length
1911: contig of 891 bp in length
2801: gap of unknown length
3779: contig of 878 bp in length
3780: gap of unknown length
4737: contig of 858 bp in length
4738: gap of unknown length
5722: contig of 885 bp in length
5723: gap of unknown length
5823: contig of 881 bp in length
6704: gap of unknown length
7684: contig of 881 bp in length
7685: gap of unknown length
7785: contig of 892 bp in length
8677: gap of unknown length
8678: contig of 896 bp in length
9673: gap of unknown length
9773: contig of 883 bp in length
10655: gap of unknown length
10755: gap of unknown length
11654: contig of 909 bp in length
11655: gap of unknown length
11765: contig of 893 bp in length
12657: gap of unknown length
12757: contig of 873 bp in length
13630: gap of unknown length
13730: gap of unknown length
14598: contig of 868 bp in length
14598: gap of unknown length
14699: contig of 878 bp in length
15576: gap of unknown length
15577: contig of 873 bp in length
16549: gap of unknown length
16550: contig of 922 bp in length
17571: gap of unknown length
17572: contig of 872 bp in length
18543: gap of unknown length
18544: contig of 879 bp in length
19522: gap of unknown length
19523: contig of 898 bp in length
20520: gap of unknown length
20521: gap of unknown length
21523: contig of 903 bp in length
21524: gap of unknown length
21623: contig of 879 bp in length
22502: gap of unknown length
22503: contig of 883 bp in length
23485: gap of unknown length
23486: gap of unknown length
24468: contig of 883 bp in length
24568: gap of unknown length
25447: contig of 879 bp in length
25448: gap of unknown length
26431: contig of 884 bp in length
26531: gap of unknown length
27406: contig of 875 bp in length
27507: gap of unknown length
28395: contig of 889 bp in length
28396: gap of unknown length
28496: contig of 885 bp in length
29380: gap of unknown length
29480: contig of 899 bp in length
30379: gap of unknown length
30380: contig of 607 bp in length
31086: contig of 607 bp in length
31186: gap of unknown length
32077: contig of 891 bp in length
32177: gap of unknown length
33054: contig of 877 bp in length
33055: gap of unknown length
34034: contig of 880 bp in length
34035: gap of unknown length
35023: contig of 889 bp in length
35123: gap of unknown length
36014: contig of 891 bp in length
36114: gap of unknown length
37015: contig of 901 bp in length
37115: gap of unknown length
37998: contig of 883 bp in length
38098: gap of unknown length
38973: contig of 875 bp in length
38974: gap of unknown length
39073: contig of 870 bp in length
40043: gap of unknown length
40044: contig of 876 bp in length
41019: gap of unknown length
41019: contig of 897 bp in length
42016: gap of unknown length
42926: contig of 910 bp in length
43026: gap of unknown length
43027: contig of 873 bp in length
43999: gap of unknown length
44000: contig of 903 bp in length
45002: gap of unknown length
45879: contig of 877 bp in length
45880: gap of unknown length
46859: contig of 880 bp in length
46860: gap of unknown length
47863: contig of 904 bp in length

```







Query Match 0.6%; Score 32; DB 80; Length 283512;  
Best Local Similarity 100.0%; Pred. No. 2.1e-05;

\* be preserved.  
 \* 1 799: contig of 799 bp in length  
 \* 800 899: gap of 100 bp



•



\* as soon as it is available and the accession number will  
\* be preserved.

1 1063: contig of 1063 bp in length  
\* 1064: gap of unknown length  
\* 1164: contig of 1055 bp in length  
\* 2219: gap of unknown length  
\* 2319: contig of 1123 bp in length  
\* 3441: gap of unknown length  
\* 3541: contig of 1122 bp in length  
\* 4664: gap of unknown length  
\* 4764: contig of 1119 bp in length  
\* 5882: gap of unknown length  
\* 5983: contig of 1042 bp in length  
\* 7025: gap of unknown length  
\* 7125: contig of 1451 bp in length  
\* 8576: gap of unknown length  
\* 8676: contig of 1129 bp in length  
\* 9805: gap of unknown length  
\* 9805: contig of 1712 bp in length  
\* 11617: gap of unknown length  
\* 11717: contig of 1007 bp in length  
\* 12724: gap of unknown length  
\* 12824: contig of 1696 bp in length  
\* 14520: gap of unknown length  
\* 14520: contig of 1095 bp in length  
\* 15715: gap of unknown length  
\* 15715: contig of 1179 bp in length  
\* 16994: gap of unknown length  
\* 17094: contig of 1128 bp in length  
\* 18222: gap of unknown length  
\* 18222: contig of 1082 bp in length  
\* 19404: gap of unknown length  
\* 20524: contig of 1021 bp in length  
\* 20525: gap of unknown length  
\* 20625: contig of 1594 bp in length  
\* 22319: gap of unknown length  
\* 22319: contig of 1049 bp in length  
\* 23368: gap of unknown length  
\* 23468: contig of 1076 bp in length  
\* 24544: gap of unknown length  
\* 24644: contig of 1071 bp in length  
\* 25715: gap of unknown length  
\* 25815: contig of 1502 bp in length  
\* 27317: gap of unknown length  
\* 27417: contig of 1104 bp in length  
\* 28520: gap of unknown length  
\* 28621: contig of 1558 bp in length  
\* 30179: gap of unknown length  
\* 30279: contig of 1200 bp in length  
\* 31479: gap of unknown length  
\* 31479: contig of 1058 bp in length  
\* 32637: gap of unknown length  
\* 32737: contig of 1045 bp in length  
\* 33782: gap of unknown length  
\* 33882: contig of 1425 bp in length  
\* 35307: gap of unknown length  
\* 35406: contig of 1119 bp in length  
\* 36525: gap of unknown length  
\* 36526: contig of 1301 bp in length  
\* 37927: gap of unknown length  
\* 38026: contig of 1300 bp in length  
\* 39327: gap of unknown length  
\* 39327: contig of 1751 bp in length  
\* 41178: gap of unknown length  
\* 41278: contig of 2268 bp in length  
\* 43545: gap of unknown length  
\* 43546: contig of 1081 bp in length  
\* 43646: gap of unknown length  
\* 44727: contig of 1267 bp in length  
\* 44826: gap of unknown length  
\* 46033: contig of 1559 bp in length  
\* 46193: gap of unknown length  
\* 46194: contig of 1559 bp in length  
\* 47753: gap of unknown length  
\* 47852: contig of 1776 bp in length  
\* 49628: contig of 1776 bp in length

49629: gap of unknown length  
\* 49729: contig of 1089 bp in length  
\* 50818: gap of unknown length  
\* 50918: contig of 1133 bp in length  
\* 52051: gap of unknown length  
\* 52151: contig of 1851 bp in length  
\* 54002: gap of unknown length  
\* 54102: contig of 1311 bp in length  
\* 55413: gap of unknown length  
\* 55513: contig of 1557 bp in length  
\* 57070: gap of unknown length  
\* 57170: contig of 1296 bp in length  
\* 58466: gap of unknown length  
\* 58566: contig of 1846 bp in length  
\* 60412: gap of unknown length  
\* 60512: contig of 1731 bp in length  
\* 62343: gap of unknown length  
\* 62343: contig of 2121 bp in length  
\* 64564: gap of unknown length  
\* 64564: contig of 2617 bp in length  
\* 67181: gap of unknown length  
\* 67281: contig of 2004 bp in length  
\* 69285: gap of unknown length  
\* 69385: contig of 1925 bp in length  
\* 71310: gap of unknown length  
\* 71410: contig of 3209 bp in length  
\* 74619: gap of unknown length  
\* 74719: contig of 2727 bp in length  
\* 77446: gap of unknown length  
\* 77546: contig of 3305 bp in length  
\* 80851: gap of unknown length  
\* 80951: contig of 3176 bp in length  
\* 84127: gap of unknown length  
\* 84227: contig of 3811 bp in length  
\* 88038: gap of unknown length  
\* 88138: contig of 2680 bp in length  
\* 90818: gap of unknown length  
\* 90918: contig of 3534 bp in length  
\* 94452: gap of unknown length  
\* 94552: contig of 3087 bp in length  
\* 97639: gap of unknown length  
\* 97739: contig of 2745 bp in length  
\* 100484: gap of unknown length  
\* 100584: contig of 2489 bp in length  
\* 103073: gap of unknown length  
\* 103173: contig of 4040 bp in length  
\* 107213: gap of unknown length  
\* 107313: contig of 3536 bp in length  
\* 110849: gap of unknown length  
\* 110949: contig of 2883 bp in length  
\* 113832: gap of unknown length  
\* 113932: contig of 3996 bp in length  
\* 117928: gap of unknown length  
\* 118028: contig of 5035 bp in length  
\* 123063: gap of unknown length  
\* 123163: contig of 4361 bp in length  
\* 127524: gap of unknown length  
\* 127624: contig of 5132 bp in length  
\* 132756: gap of unknown length  
\* 132856: contig of 8828 bp in length  
\* 141684: gap of unknown length  
\* 141784: contig of 7554 bp in length  
\* 149338: gap of unknown length  
\* 149438: contig of 10601 bp in length  
\* 160039: gap of unknown length  
\* 160139: contig of 12150 bp in length  
\* 172289: gap of unknown length  
\* 172389: contig of 19487 bp in length  
\* 191876: gap of unknown length  
\* 191976: contig of 18819 bp in length  
\* 210795: gap of unknown length  
\* 210895: contig of 28192 bp in length

Location/Qualifiers

FEATURES











numbers given in the feature table with their source databases:  
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr9>  
 RP11-509J21 is from the library RPCI-11.2 constructed at the  
 Roswell Park Cancer Institute by the group of Pieter de Jong. For  
 further details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-509J21. The true  
 left end of clone RP11-661K19 is at 114307 in this sequence. The  
 true right end of clone RP11-326A8 is at 32194 in this sequence.

## FEATURES

source	Location/Qualifiers
	1..163447
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-509J21"
	/clone_lib="RPCI-11.2"
	1..547
repeat_region	/note="L1M4 repeat: matches 4282..4882 of consensus"
repeat_region	1664..1739
repeat_region	/note="AluJ/FLAM repeat: matches 6..81 of consensus"
repeat_region	1819..1872
repeat_region	/note="Alu repeat: matches 243..296 of consensus"
repeat_region	2942..3214
repeat_region	/note="AluSx repeat: matches 39..312 of consensus"
repeat_region	3581..3702
repeat_region	/note="AluJ/FRAM repeat: matches 185..295 of consensus"
repeat_region	3999..4195
repeat_region	/note="MIR repeat: matches 45..262 of consensus"
repeat_region	4258..4550
repeat_region	/note="AluJ repeat: matches 1..293 of consensus"
repeat_region	4797..5012
repeat_region	/note="AluSg/x repeat: matches 86..300 of consensus"
repeat_region	5026..5558
repeat_region	/note="L2 repeat: matches 1710..2278 of consensus"
repeat_region	5978..6102
repeat_region	/note="L2 repeat: matches 2345..2474 of consensus"
repeat_region	6113..6349
repeat_region	/note="MIR repeat: matches 11..259 of consensus"
repeat_region	6560..6861
repeat_region	/note="AluSg repeat: matches 1..302 of consensus"
repeat_region	8760..8975
repeat_region	/note="AluJb repeat: matches 72..287 of consensus"
repeat_region	9395..9659
repeat_region	/note="AluSx repeat: matches 1..284 of consensus"
repeat_region	10786..11196
repeat_region	/note="MLT1J repeat: matches 1..414 of consensus"
repeat_region	11402..11528
repeat_region	/note="MIR repeat: matches 20..149 of consensus"
repeat_region	11632..11659
repeat_region	/note="14 copies 2 mer ac 100% conserved"
repeat_region	11729..11921
repeat_region	/note="MIR repeat: matches 31..221 of consensus"
repeat_region	11943..12062
repeat_region	/note="L1ME3A repeat: matches 5991..6113 of consensus"
repeat_region	12063..12621
repeat_region	/note="MLT2D repeat: matches 1..553 of consensus"
repeat_region	12622..12672
repeat_region	/note="L1ME3A repeat: matches 6113..6164 of consensus"
repeat_region	12742..13045
repeat_region	/note="AluSg repeat: matches 1..305 of consensus"
repeat_region	13420..13511
repeat_region	/note="L1MC2 repeat: matches 6218..6314 of consensus"
repeat_region	13692..20142
repeat_region	/note="L1PA7 repeat: matches 8..6139 of consensus"
repeat_region	20151..20339
repeat_region	/note="L1MC2 repeat: matches 6044..6234 of consensus"
repeat_region	20353..20841
	/note="Tigger3b repeat: matches 701..1205 of consensus"
	20842..21140
	/note="AluJb repeat: matches 1..288 of consensus"
	21141..21896
	/note="Tigger3b repeat: matches 1..701 of consensus"
	21915..22265
	/note="L1MC2 repeat: matches 5643..6004 of consensus"
	22435..22736
	/note="AluSg repeat: matches 1..303 of consensus"
	23083..23379
	/note="AluSx repeat: matches 3..297 of consensus"
	23984..24170
	/note="MIR repeat: matches 49..244 of consensus"
	25343..25466
	/note="AluSg/x repeat: matches 170..293 of consensus"
	25921..26189
	/note="MER58C repeat: matches 4..89 of consensus"
	27058..27825
	/note="L1MB3 repeat: matches 5446..6176 of consensus"
	27881..27928
	/note="MIR repeat: matches 49..96 of consensus"
	28125..28584
	/note="L2 repeat: matches 1992..2489 of consensus"
	28661..28889
	/note="L2 repeat: matches 1643..1880 of consensus"
	29272..29429
	/note="MER5B repeat: matches 1..175 of consensus"
	29485..29731
	/note="MIR repeat: matches 3..262 of consensus"
	31630..32055
	/note="L2 repeat: matches 2099..2545 of consensus"
	32445..32490
	/note="23 copies 2 mer aa 73% conserved"
	32838..33074
	/note="MER20 repeat: matches 2..218 of consensus"
	33177..33580
	/note="MLT1B repeat: matches 1..390 of consensus"
	34501..34800
	/note="AluJ repeat: matches 1..299 of consensus"
	34910..35029
	/note="L2 repeat: matches 1931..2053 of consensus"
	35051..36193
	/note="MER69B repeat: matches 1..1225 of consensus"
	36596..36913
	/note="L1MD repeat: matches 977..1287 of consensus"
	37179..37473
	/note="AluJb repeat: matches 1..307 of consensus"
	37497..37538
	/note="21 copies 2 mer aa 83% conserved"
	37563..37976
	/note="MSTA repeat: matches 1..426 of consensus"
	37977..38328
	/note="MSTA repeat: matches 1..426 of consensus"
	38341..38696
	/note="L1MEC repeat: matches 1639..1982 of consensus"
	38762..38914
	/note="AluJ/FRAM repeat: matches 150..302 of consensus"
	39048..40267
	/note="L1MEC repeat: matches 2189..3119 of consensus"
	40296..41553
	/note="L2 repeat: matches 433..1770 of consensus"
	41554..41851
	/note="AluJb repeat: matches 1..312 of consensus"
	41852..42145
	/note="L2 repeat: matches 1770..2050 of consensus"
	43119..43335
	/note="MIR repeat: matches 6..262 of consensus"
	43646..44014
	/note="MSTA repeat: matches 1..375 of consensus"
	44015..44395
	/note="THE1C repeat: matches 1..371 of consensus"
	44396..44448
	/note="MSTA repeat: matches 375..426 of consensus"



















Query Match 0.8%; Score 41; DB 80; Length 195380;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 tcatatacatatgtttgagttagtagaatacaatgaat 523  
 |||||  
 DB 170711 TCATATACATATGCTTTGAGTTAGTAGAATGAAT 170751

RESULT 24  
 AC007038 99030 bp DNA PRI 30-SEP-2000  
 LOCUS Homo sapiens BAC clone RP11-260M2 from 2, complete sequence.  
 DEFINITION AC007038  
 ACCESSION AC007038  
 VERSION AC007038.3 GI:5708475  
 KEYWORDS HTG  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 99030)  
 Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 99030)  
 COURTNEY, L., WOHLDMANN, P. and KETTERMAN, M.  
 TITLE The sequence of Homo sapiens BAC clone RP11-260M2  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 99030)  
 WATERSTON, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-AUG-1999) Genome Sequencing Center, Washington  
 UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 4 (bases 1 to 99030)  
 WATERSTON, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-AUG-1999) Genome Sequencing Center, Washington  
 UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 99030)  
 WATERSTON, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Aug 8, 1999 this sequence version replaced gi:4580502.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0260M02  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30; an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this

sequence, see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (http://bacpac.med.buffalo.edu)

VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-436C12, 200 bp overlap; the  
 clone sequenced to the right is RP11-396J78. Actual start of this  
 clone is at base position 88990 of RP11-436C12.

#### FEATURES

Location/Qualifiers	source
1..99030	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="2"	
/map="2"	
/clone_lib="RPC1-11"	
/clone="RP11-260M2"	
195..483	repeat_region
/rpt_family="MaLR"	
484..549	repeat_region
/rpt_family="L1"	
550..665	repeat_region
/rpt_family="MaLR"	
1489..1530	repeat_region
/rpt_family="(TTG)n"	
2391..2427	repeat_region
/rpt_family="AT-rich"	
4953..4978	repeat_region
/rpt_family="(TG)n"	
4978..5005	repeat_region
/rpt_family="(GA)n"	
5211..5320	repeat_region
/rpt_family="L2"	
5653..6295	repeat_region
/rpt_family="MER21-group"	
7089..7196	repeat_region
/rpt_family="CT-rich"	
7939..7986	repeat_region
/rpt_family="MER1_type"	
8042..8397	repeat_region
/rpt_family="L1"	
8398..8693	repeat_region
/rpt_family="Alu"	
8694..9033	repeat_region
/rpt_family="L1"	
9045..9232	repeat_region
/rpt_family="Alu"	
9356..9611	repeat_region
/rpt_family="Alu"	
10944..11018	repeat_region
/rpt_family="(TATATG)n"	
11295..11326	repeat_region
/rpt_family="AT-rich"	
12027..12521	misc_feature
/note="match to EST A1791679 (NID:g5339395) op98c06.y5"	
12079..12639	misc_feature
/note="match to EST AA939008 (NID:g3098921) op98c06.s1"	
15316..15343	repeat_region
/rpt_family="AT-rich"	
17192..17229	repeat_region
/rpt_family="(TATATG)n"	
17230..17258	repeat_region
/rpt_family="(TAGA)n"	
18462..18694	repeat_region
/rpt_family="MIR"	
18809..18837	repeat_region



misc\_feature /note="assembly\_fragment:01062  
fragment\_chain:1"  
2432..5660  
misc\_feature /note="assembly\_fragment:01573  
fragment\_chain:1"  
5761..8322  
misc\_feature /note="assembly\_fragment:00193  
fragment\_chain:1"  
8423..13132  
misc\_feature /note="assembly\_fragment:00386  
fragment\_chain:1"  
13233..20498  
misc\_feature /note="assembly\_fragment:01407  
fragment\_chain:2"  
20599..25933  
misc\_feature /note="assembly\_fragment:00948  
fragment\_chain:2"  
26034..31897  
misc\_feature /note="assembly\_fragment:00950  
fragment\_chain:3"  
31998..43682  
misc\_feature /note="assembly\_fragment:01201  
fragment\_chain:2"  
43783..55739  
misc\_feature /note="assembly\_fragment:00950  
fragment\_chain:3"  
55840..62497  
misc\_feature /note="assembly\_fragment:00010  
fragment\_chain:3"  
62598..66053  
misc\_feature /note="assembly\_fragment:00516  
fragment\_chain:3"  
66154..71233  
misc\_feature /note="assembly\_fragment:00799  
fragment\_chain:4"  
71334..78676  
misc\_feature /note="assembly\_fragment:01288  
fragment\_chain:4"  
78777..83031  
misc\_feature /note="assembly\_fragment:01254  
fragment\_chain:5"  
83132..87164  
misc\_feature /note="assembly\_fragment:01420  
fragment\_chain:5"  
87265..96445  
misc\_feature /note="assembly\_fragment:00385"  
96546..99250  
misc\_feature /note="assembly\_fragment:00401"  
99351..109824  
misc\_feature /note="assembly\_fragment:00662"  
109925..114821  
misc\_feature /note="assembly\_fragment:00921"  
114922..117686  
misc\_feature /note="assembly\_fragment:01024"  
117787..123506  
misc\_feature /note="assembly\_fragment:01232"  
123607..142472  
misc\_feature /note="assembly\_fragment:01449"  
142573..151055  
misc\_feature /note="assembly\_fragment:01538"  
151156..153665  
misc\_feature /note="assembly\_fragment:01612"  
29528 c 29581 g 45550 t 2314 others

BASE COUNT 46692 a 29528 c 29581 g 45550 t 2314 others  
ORIGIN

Query Match 0.9%; Score 46; DB 78; Length 153665;  
Best Local Similarity 100.0%; Pred. No. 5.3e-13;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
45 aggggtagaataattctgtcattgcattcaagactaggacca 90  
|||||  
DB 151241 AGGGGTAGAATAATTTCTGTCTGCTCATTCAAAGACTAGGACCA 151196

RESULT 23  
AL358892  
LOCUS  
DEFINITION  
PROGRESS \*\*\*, in unordered pieces.  
ACCESSION  
AL358892  
VERSION  
GI:11493289  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 195380)  
AUTHORS  
Wall, M.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
COMMENT  
On Nov 30, 2000 this sequence version replaced gi:11414592.

AL358892 195380 bp DNA HTG 06-MAR-2001  
Mus musculus chromosome 5 clone RP21-417G6, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in unordered pieces.  
AL358892  
AL358892.12 GI:11493289  
HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
house mouse.  
Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 195380)  
Direct Submission  
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Nov 30, 2000 this sequence version replaced gi:11414592.

Center: Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: http://mrcseq.har.mrc.ac.uk  
Contact: mouseseq@har.mrc.ac.uk  
----- Project Information  
Center project name: dm417G6  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 190612 bases at least Q40  
Consensus quality: 192224 bases at least Q30  
Consensus quality: 193166 bases at least Q20  
Insert size: 194680; sum-of-contigs  
Insert size: 191947; 3.7% error; agarose-fp  
Quality coverage: 7.33x in Q20 bases; sum-of-contigs Quality  
coverage: 8.85x in Q20 bases; agarose-fp

\*\*\*\*\*  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
Source Location/Qualifiers  
1..195380  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/clone="RP21-417G6"  
/clone\_lib="RPCI-21"  
1..2333  
/note="assembly\_fragment:02826  
fragment\_chain:1"  
2434..4472  
/note="assembly\_fragment:00159  
fragment\_chain:1"  
4573..6623  
/note="assembly\_fragment:00292"  
6724..8757  
/note="assembly\_fragment:02097"  
8858..11498  
/note="assembly\_fragment:02560"  
11599..13673  
/note="assembly\_fragment:03440"  
13774..16148  
/note="assembly\_fragment:03977"  
16249..195380  
/note="assembly\_fragment:04183"  
55503 a 40458 c 40664 g 58048 t 707 others  
BASE COUNT  
ORIGIN



```

/note="assembly_fragment:00516
fragment_chain:3"
misc_feature
66154..71233
/note="assembly_fragment:00799
fragment_chain:4"
misc_feature
71334..78676
/note="assembly_fragment:01288
fragment_chain:4"
misc_feature
78777..83031
/note="assembly_fragment:01254
fragment_chain:5"
misc_feature
83132..87164
/note="assembly_fragment:01420
fragment_chain:5"
misc_feature
87265..96445
/note="assembly_fragment:00385"
misc_feature
96546..99250
/note="assembly_fragment:00401"
misc_feature
99351..109824
/note="assembly_fragment:00662"
misc_feature
109925..114821
/note="assembly_fragment:00921"
misc_feature
114922..117686
/note="assembly_fragment:01024"
misc_feature
117787..123506
/note="assembly_fragment:01232"
misc_feature
123607..142472
/note="assembly_fragment:01449"
misc_feature
142573..151055
/note="assembly_fragment:01538"
misc_feature
151156..153665
/note="assembly_fragment:01612"

```

BASE COUNT 46692 a 29528 c 29581 g 45550 t 2314 others  
ORIGIN

Query Match 0.9%; Score 47; DB 78; Length 153665;  
Best Local Similarity 100.0%; Pred. No. 1.5e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ccggagagcccgagtagcgagtagcgagtcgcgaaccgcggagg 47
|||||
DB 62250 CCGGAGAGCCCGGAGTAGCGGAGTAGCGGAGTCGCAACCCGGAGG 62296

```

```

RESULT 22
AL138818/c
LOCUS      AL138818      153665 bp      DNA      HTG      18-AUG-2000
DEFINITION Homo sapiens chromosome 13 clone RP11-345N9, *** SEQUENCING IN
PROGRESS ***, 24 unordered pieces.
ACCESSION AL138818
VERSION    AL138818.5  GI:9863508
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 153665)
AUTHORS    Burton.J.
TITLE      Direct Submission
JOURNAL    Submitted (17-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9212026.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba345N9
----- Summary Statistics
Assembly program: XGAP4; version 4.5

```

Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 140298 bases at least Q40  
Consensus quality: 144322 bases at least Q30  
Consensus quality: 147003 bases at least Q20  
Insert size: 151365; sum-of-contigs  
Quality coverage: 3.63x in Q20 bases; agarose-fp  
coverage: 3.40x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1
* 2332 2431: gap of 100 bp
* 2432 5660: contig of 3229 bp in length
* 5661 5760: gap of 100 bp
* 5761 8322: contig of 2562 bp in length
* 8323 8422: gap of 100 bp
* 8423 13132: contig of 4710 bp in length
* 13133 13232: gap of 100 bp
* 13233 20498: contig of 7266 bp in length
* 20499 20598: gap of 100 bp
* 20599 25933: contig of 5335 bp in length
* 25934 26033: gap of 100 bp
* 26034 31897: contig of 5864 bp in length
* 31898 31997: gap of 100 bp
* 31998 43682: contig of 11685 bp in length
* 43683 43782: gap of 100 bp
* 43783 55739: contig of 11957 bp in length
* 55740 55839: gap of 100 bp
* 55840 62497: contig of 6658 bp in length
* 62498 62597: gap of 100 bp
* 62598 66053: contig of 3456 bp in length
* 66054 66153: gap of 100 bp
* 66154 71233: contig of 5080 bp in length
* 71234 71333: gap of 100 bp
* 71334 78676: contig of 7343 bp in length
* 78677 78776: gap of 100 bp
* 78777 83031: contig of 4255 bp in length
* 83032 83131: gap of 100 bp
* 83132 87164: contig of 4033 bp in length
* 87165 87264: gap of 100 bp
* 87265 96445: contig of 9181 bp in length
* 96446 96545: gap of 100 bp
* 96546 99250: contig of 2705 bp in length
* 99251 99350: gap of 100 bp
* 99351 109824: contig of 10474 bp in length
* 109825 109924: gap of 100 bp
* 109925 114821: contig of 4897 bp in length
* 114822 114921: gap of 100 bp
* 114922 117686: contig of 2765 bp in length
* 117687 117786: gap of 100 bp
* 117787 123506: contig of 5720 bp in length
* 123507 123606: gap of 100 bp
* 123607 142472: contig of 18866 bp in length
* 142473 142572: gap of 100 bp
* 142573 151055: contig of 8483 bp in length
* 151056 151155: gap of 100 bp
* 151156 153665: contig of 2510 bp in length.

```

Location/Qualifiers  
1..153665  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-345N9"  
/clone\_lib="RPC1-11.2"  
1..2331  
misc\_feature







repeat\_region /note="L1MB4 repeat: matches 2. .91 of consensus"  
817. .1109  
repeat\_region /note="AluSg repeat: matches 299. .5 of consensus"  
1113. .1165  
repeat\_region /note="L1MA4 repeat: matches 87. .138 of consensus"  
1195. .1249  
repeat\_region /note="AluJ repeat: matches 1. .55 of consensus; incomplete  
1352. .1551  
repeat\_region /note="AluSg repeat: matches 3. .303 of consensus"  
1593. .1797  
repeat\_region /note="MSTB repeat: matches 185. .391 of consensus"  
1669. .1796  
repeat\_region /note="MSTB repeat: matches 289. .422 of consensus"  
3346. .3416  
repeat\_region /note="L1MB4 repeat: matches 859. .929 of consensus"  
3418. .3552  
repeat\_region /note="L1 repeat: matches 4982. .5119 of consensus"  
3558. .3838  
repeat\_region /note="AluSg repeat: matches 1. .282 of consensus;  
incomplete repeat"  
4035. .4117  
repeat\_region /note="L1MB3 repeat: matches 5. .76 of consensus"  
4120. .4429  
repeat\_region /note="AluJb repeat: matches 3. .302 of consensus"  
4561. .5055  
repeat\_region /note="L1MB8 repeat: matches 69. .572 of consensus"  
5056. .5186  
repeat\_region /note="FLAM\_C repeat: matches 1. .132 of consensus"  
5188. .5500  
repeat\_region /note="L1MB3 repeat: matches 567. .923 of consensus"  
6890. .7101  
repeat\_region /note="MIR repeat: matches 15. .228 of consensus"  
7430. .7715  
repeat\_region /note="AluJb repeat: matches 1. .292 of consensus"  
7718. .8028  
repeat\_region /note="AluX repeat: matches 1. .302 of consensus"  
8460. .8763  
repeat\_region /note="AluSg repeat: matches 302. .1 of consensus"  
8979. .9338  
repeat\_region /note="L1MA1 repeat: matches 1. .365 of consensus"  
9635. .9799  
repeat\_region /note="L1MA5A repeat: matches 636. .472 of consensus"  
9670. .9814  
repeat\_region /note="L1PA15 repeat: matches 600. .458 of consensus"  
11277. .11857  
repeat\_region /note="L1PA15 repeat: matches 311. .902 of consensus"  
12059. .12316  
repeat\_region /note="THELC repeat: matches 371. .121 of consensus"  
12476. .12670  
repeat\_region /note="AluJo repeat: matches 301. .107 of consensus;  
incomplete repeat"  
12673. .12959  
repeat\_region /note="AluSg repeat: matches 286. .1 of consensus"  
12968. .13106  
repeat\_region /note="FLAM\_C repeat: matches 133. .2 of consensus"  
13148. .13425  
repeat\_region /note="AluSg repeat: matches 1. .300 of consensus"  
13562. .14154  
repeat\_region /note="L1MB1 repeat: matches 610. .3 of consensus"  
14009. .14231  
repeat\_region /note="L1 repeat: matches 5390. .5167 of consensus"  
14258. .14414  
repeat\_region /note="L1MA4 repeat: matches 742. .905 of consensus"  
14483. .14608  
repeat\_region /note="FLAM\_A repeat: matches 1. .126 of consensus"  
14620. .14736  
repeat\_region /note="L1MA4A repeat: matches 930. .1042 of consensus"  
15223. .15816  
repeat\_region /note="AluSc repeat: matches 294. .1 of consensus"  
15867. .16161  
repeat\_region /note="AluX repeat: matches 1. .293 of consensus"  
16162. .16364

repeat\_region /note="L1 repeat: matches 4313. .4112 of consensus"  
16668. .16953  
repeat\_region /note="AluJ repeat: matches 1. .295 of consensus"  
17125. .17538  
prim\_transcript /note="match: multiple ESTs; low % ID; match: T596663  
R48957 H24241 CL4861 F03338; match: W88836 N22360 H70851  
F02679 N35317; match: R10942 N73317 N20948 R36018 R42572;  
match: R62472 R43248 R41198 R38482 R42487; match: R05993  
N71787 T64627 F03312"  
18303. .18654  
repeat\_region /note="L1MA1 repeat: matches 1. .357 of consensus"  
18784. .19084  
repeat\_region /note="AluX repeat: matches 2. .302 of consensus"  
20018. .20312  
repeat\_region /note="AluSg repeat: matches 298. .2 of consensus"  
21304. .21583  
repeat\_region /note="AluX repeat: matches 289. .1 of consensus"  
22222. .22251  
repeat\_region /note="15 copies of 2 mer 90 % conserved"  
22501. .22555  
repeat\_region /note="THELC repeat: matches 371. .316 of consensus"  
22624. .22950  
repeat\_region /note="MSTA repeat: matches 372. .1 of consensus"  
23311. .23601  
repeat\_region /note="AluSg repeat: matches 299. .3 of consensus"  
23864. .24165  
repeat\_region /note="AluSp repeat: matches 303. .1 of consensus"  
24267. .24562  
repeat\_region /note="AluSc repeat: matches 4. .299 of consensus"  
24559. .24604  
unsure /note="single clone"  
24931. .25227  
repeat\_region /note="AluSg repeat: matches 294. .1 of consensus"  
25424. .25562  
repeat\_region /note="AluSg repeat: matches 139. .1 of consensus;  
incomplete repeat"  
25564. .26134  
repeat\_region /note="L1MA5 repeat: matches 5390. .5077 of consensus"  
25987. .26296  
repeat\_region /note="L1 repeat: matches 426. .2 of consensus"  
26318. .26717  
repeat\_region /note="MSTA repeat: matches 426. .2 of consensus"  
26837. .27018  
repeat\_region /note="AluX repeat: matches 302. .116 of consensus;  
incomplete repeat"  
27019. .27146  
repeat\_region /note="L1 repeat: matches 5022. .4893 of consensus"  
27316. .27534  
repeat\_region /note="AluJb repeat: matches 302. .85 of consensus;  
incomplete repeat"  
27720. .27880  
repeat\_region /note="MSTB repeat: matches 157. .2 of consensus"  
28372. .28672  
repeat\_region /note="AluSp repeat: matches 303. .2 of consensus"  
28936. .29047  
repeat\_region /note="MIR2 repeat: matches 138. .24 of consensus"  
29097. .29475  
repeat\_region /note="L1MA2 repeat: matches 1. .365 of consensus"  
29474. .29670  
repeat\_region /note="MER2 repeat: matches 4. .206 of consensus"  
29751. .29797  
repeat\_region /note="L1MA2 repeat: matches 507. .553 of consensus"  
30487. .30726  
repeat\_region /note="AluJo repeat: matches 239. .1 of consensus;  
incomplete repeat"  
31154. .31451  
repeat\_region /note="AluSc repeat: matches 298. .1 of consensus"  
31628. .32166  
repeat\_region /note="L1 repeat: matches 4854. .5390 of consensus"  
32017. .32906  
repeat\_region /note="L1PA2 repeat: matches 1. .891 of consensus"  
32909. .32938  
repeat\_region /note="15 copies of 2 mer 87 % conserved"



```

/note="assembly_fragment:03653
fragment_chain:5"
misc_feature 158093..164601
/note="assembly_fragment:05568
fragment_chain:5"
misc_feature 164702..195901
/note="assembly_fragment:01977
fragment_chain:5"
misc_feature 196002..198807
/note="assembly_fragment:00023
fragment_chain:6"
misc_feature 198908..204604
/note="assembly_fragment:05254
fragment_chain:6"
misc_feature 204705..207838
/note="assembly_fragment:00245
fragment_chain:7"
misc_feature 207939..221335
/note="assembly_fragment:03587
fragment_chain:7"
misc_feature 221436..224577
/note="assembly_fragment:00642
fragment_chain:8"
misc_feature 224678..228248
/note="assembly_fragment:04656
fragment_chain:8"
misc_feature 228349..237923
/note="assembly_fragment:01666
fragment_chain:9"
misc_feature 238024..242134
/note="assembly_fragment:05529
fragment_chain:9"
misc_feature 242235..250341
/note="assembly_fragment:02430
fragment_chain:10"
misc_feature 250442..273524
/note="assembly_fragment:00629
fragment_chain:10"
misc_feature 273625..277125
/note="assembly_fragment:04232
fragment_chain:11"
misc_feature 277226..282694
/note="assembly_fragment:02725
fragment_chain:11"
misc_feature 282795..284888
/note="assembly_fragment:04893
fragment_chain:12"
misc_feature 284989..290372
/note="assembly_fragment:01882
fragment_chain:12"
misc_feature 290473..298809
/note="assembly_fragment:00742"
misc_feature 298910..301400
/note="assembly_fragment:00762"
misc_feature 301501..303502
/note="assembly_fragment:01554"
misc_feature 303603..306492
/note="assembly_fragment:02456"
misc_feature 306593..309852
/note="assembly_fragment:03413"
misc_feature 309953..312398
/note="assembly_fragment:03967"
misc_feature 312499..315866
/note="assembly_fragment:04102"
misc_feature 315967..318753
/note="assembly_fragment:04448"
misc_feature 318854..323190
/note="assembly_fragment:04616"
misc_feature 323291..329923
/note="assembly_fragment:04963"
misc_feature 330024..332405
/note="assembly_fragment:06188"
misc_feature 332506..335631

```

```

misc_feature /note="assembly_fragment:06207"
335732..338304
misc_feature /note="assembly_fragment:06594"
338405..340969
/note="assembly_fragment:06804"
BASE COUNT 100857 a 71110 c 68176 g 96314 t 4512 others
ORIGIN

Query Match 1.1%; Score 56; DB 81; Length 340969;
Best Local Similarity 100.0%; Pred. No. 2.3e-18;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3006 gaaataattgtctctaccagagatattgtccatacaattcaccctttggca 3065
|||||
Db 61173 GAAAAATTATGTCTCTTCTACCAGATATGTTGTTCCATACAAATTCACCTTTGGCA 61114
|||||

QY 3066 catgaccagattatgtcaagtagcagagatattgaacaaacttaagatgt 3115
|||||
Db 61113 CATGACCAGATTATGTCAAAGTACAGATATTGAACAACCTTAAGATGT 61064
|||||

RESULT 20
HS26H23 91835 bp DNA PRI 22-NOV-1999
LOCUS Human DNA sequence from PAC 26H23, BRCA2 gene region chromosome
DEFINITION 13q12-13 contains ESTs, CpG island.
ACCESSION 284467.1 GI:2104578
VERSION 284467.1
KEYWORDS 13q12-13; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91835)
AUTHORS Williamson, H.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
ON May 17, 1997 this sequence version replaced gi:1806009.
IMPORTANT:
This sequence is not the entire insert of clone 26H23. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 26H23 is at 1 in this sequence. The true
right end of clone 130N4 is at 34585.
The true left end of clone 267P19 is at 91732.
26H23 is from the human PAC library described in Ioannou A.P. et al
Nature Genet 6, 84-89.
FEATURES
source
1. 91835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-13"
/clone_lib="RPC1-1"
/clone="XX-26H23"
1. 83
/note="AluSc repeat: matches 84. .1 of consensus;
incomplete repeat"
repeat_region 94..396
/note="AluSx repeat: matches 302. .1 of consensus"
repeat_region 451..710
/note="L1 repeat: matches 4886. .5146 of consensus"
repeat_region 708..815

```



## PCR Profile:

Presoak: 94 degrees C for 4.00 minutes  
 Denaturation: 94 degrees C for 50.0 seconds  
 Annealing: 58 degrees C for 1.50 minutes  
 Polymerization: 72 degrees C for 1.00 minutes  
 PCR Cycles: 30  
 Thermal Cycler: custom built by IAS, Costar, Cambridge MA

## Protocol:

Template: 10 ng  
 Primer: each 5 pM  
 dNTPs: 4 mM  
 Taq Polymerase: 0.5 U  
 Total Vol: 20 uL

## Buffer:

Mg2+: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 Gelatin: .001 %

## FEATURES

source  
 Location/Qualifiers  
 1..150  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="24.00 cR from top of Chr13 linkage group"  
 /clone\_lib="Human Thudson EST"  
 /note="STS derived from sequences in dbEST and the  
 Unigene collection."  
 STS  
 primer\_bind 1..150  
 primer\_bind 1..20  
 complement(127..150)  
 BASE COUNT 43 a 26 c 38 g 42 t 1 others  
 ORIGIN

Query Match 1.9%; Score 99; DB 54; Length 150;

Best Local Similarity 99.3%; Pred. NO. 6.2e-42;

Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaatgggactgtgaagagtgacagtggtgaccttacttggtagccaccatacatt 4550  
 |||||  
 DB 1 CACAAAATGGGACTGTGGAAGAGTGGACAGTTGGACCTTACTTTGGTAGCCCATACATT 60  
 QY 4551 tgggtcacatgttttagccatcacatggttaacattgactatggagcttcttggaagtg 4610  
 |||||  
 DB 61 TGTGGTCACATGCTTTAGGCATACNCATGCTTAACATTGACTGGAGTCTTGTGAAAGTG 120  
 QY 4611 taatgtcgtatggtctatgtagacataaaga 4640  
 |||||  
 DB 121 TAATGTGGATGGCTATGTAGACATAAAGA 150

## RESULT 19

AL512630/c  
 LOCUS 340969 bp DNA HTG 21-MAR-2001  
 DEFINITION Mus musculus chromosome 5 clone RP21-583E8, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in unordered pieces.

AL512630

ACCESSION AL512630.2 GI:13443466

VERSION HTG; HTGS\_PHASE1.

KEYWORDS house mouse.

## SOURCE

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 340969)

AUTHORS Sime, S.

TITLE Direct Submission

JOURNAL Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonequery@sanger.ac.uk

On Mar 24, 2001 this sequence version replaced gi:12193249.

## COMMENT

----- Genome Center  
 Center: UK Medical Research Council  
 Center code: UK-MRC

Web site: http://mrcseq.har.mrc.ac.uk  
 Contact: mouse@har.mrc.ac.uk

----- Project Information

Center project name: dm583E8

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 323281 bases at least Q40

Consensus quality: 329435 bases at least Q30

Consensus quality: 333234 bases at least Q20

Insert size: 336469; sum-of-contigs

Insert size: 189900; 1.3% error; agarose-fp

Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality

coverage: 11.68x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

## FEATURES

source  
 Location/Qualifiers  
 1..340969  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /clone\_lib="RP21-583E8"  
 /clone\_lib="RPGI-21"  
 1..10854  
 /note="assembly\_fragment:06116  
 fragment\_chain:1"  
 misc\_feature 10955..16884  
 /note="assembly\_fragment:05022  
 fragment\_chain:1"  
 misc\_feature 16985..22344  
 /note="assembly\_fragment:04089  
 fragment\_chain:1"  
 misc\_feature 22445..36631  
 /note="assembly\_fragment:04629  
 fragment\_chain:1"  
 misc\_feature 36732..63941  
 /note="assembly\_fragment:02431  
 fragment\_chain:1"  
 misc\_feature 64042..76959  
 /note="assembly\_fragment:06652  
 fragment\_chain:2"  
 misc\_feature 77060..80361  
 /note="assembly\_fragment:02856  
 fragment\_chain:2"  
 misc\_feature 80462..84581  
 /note="assembly\_fragment:05282  
 fragment\_chain:2"  
 misc\_feature 84682..89169  
 /note="assembly\_fragment:01199  
 fragment\_chain:2"  
 misc\_feature 89270..94454  
 /note="assembly\_fragment:00854  
 fragment\_chain:3"  
 misc\_feature 94555..99344  
 /note="assembly\_fragment:04769  
 fragment\_chain:3"  
 misc\_feature 99445..105364  
 /note="assembly\_fragment:03592  
 fragment\_chain:3"  
 misc\_feature 105465..112857  
 /note="assembly\_fragment:02797  
 fragment\_chain:4"  
 misc\_feature 112958..116353  
 /note="assembly\_fragment:05253  
 fragment\_chain:4"  
 misc\_feature 116454..119029  
 /note="assembly\_fragment:01663  
 fragment\_chain:4"  
 misc\_feature 119130..157992



```

* 51027 51126: gap of 100 bp
* 51127 51836: contig of 710 bp in length
* 51837 51936: gap of 100 bp
* 51937 52671: contig of 735 bp in length
* 52672 52771: gap of 100 bp
* 52772 53502: contig of 731 bp in length
* 53503 53602: gap of 100 bp
* 53603 54343: contig of 741 bp in length
* 54344 54443: gap of 100 bp
* 54444 55141: contig of 698 bp in length
* 55142 55241: gap of 100 bp
* 55242 55950: contig of 709 bp in length
* 55951 56050: gap of 100 bp

Query Match 1.9%; Score 100; DB 77; Length 69740;
Best Local Similarity 100.0%; Pred. No. 3.2e-42; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 0;

QY 1821 aagttgggaaccccaacagcctcacaatcctttcttggaatgatcaagtttctcttg 1880
|||||
Db 13983 AAGTTGGCAACCCCAACAGCCTACAAATCCTTCTCTGGAATGATCAAGTTCTCTTG 14042

QY 1881 gagaggatagcactgtgcacatagataccgaatctatca 1920
|||||
Db 14043 GAGAGGATGACCTGTGCACATAGATACCGAATCTATCA 14082

RESULT 17
G43328 150 bp mRNA STS 27-JAN-1999
LOCUS
DEFINITION WIAF-1982-STS Human Thudson EST Homo sapiens STS cDNA, sequence
tagged site.
ACCESSION G43328
VERSION G43328.1 GI:4192245
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
JOURNAL MEDLINE 98248615
COMMENT Synonyms: EST397175a, EST397175
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CACAAATGGGACTGCTGAA
Primer B: TCTTATGTCTACATAGCCATCGC
STS size: 150
PCR Profile:
Presoak: 94 degrees C for 4.00 minutes
Denaturation: 94 degrees C for 50.0 seconds
Annealing: 58 degrees C for 1.50 minutes
Polymerization: 72 degrees C for 1.00 minutes
PCR Cycles: 30
Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
Template: 10 ng
Primer: each 5 pm
dNTPs: 4 mM
Tag Polymerase: 0.5 U

```

```

Total Vol: 20 uL
Buffer:
Mg2+: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Gelatin: .001 %
Location/Qualifiers
1..150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="24.00 cR from top of Chr13 linkage group"
/clone_lib="Human Thudson EST"
/note="STSs derived from sequences in dbEST and the
Unigene collection."
STS
primer_bind 1..150
primer_bind 1..20
BASE COUNT 43 a 26 c 38 g 42 t 1 others
ORIGIN

Query Match 1.9%; Score 99; DB 54; Length 150;
Best Local Similarity 99.3%; Pred. No. 6.2e-42;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaaaggagctgtgaagagtgacagttggaccttacttcttggtgacccatacatt 4550
|||||
Db 1 CACAAAATGGGACTGCTGAAGAGTGGACAGTTGACCTTACTTTGGTGACCCCATACNTT 60

QY 4551 tgtgtcacatgcttttagccatacacatggttaacattgacatgagctctgtgaaagt 4610
|||||
Db 61 TGTGTCACATGCTTTAGCCATACACATGGTAAACATGTGACTATGAGTCTTGTGAAGTG 120

QY 4611 taatgtgcgatggctatgtagacataaaga 4640
|||||
Db 121 TAATGTGCGATGGCTATGTAGACATAAAGA 150

RESULT 18
G43329 150 bp mRNA STS 27-JAN-1999
LOCUS
DEFINITION WIAF-1983-STS Human Thudson EST Homo sapiens STS cDNA, sequence
tagged site.
ACCESSION G43329
VERSION G43329.1 GI:4192246
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
JOURNAL MEDLINE 98248615
COMMENT Synonyms: EST397175b, EST397175
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CACAAATGGGACTGCTGAA
Primer B: TCTTATGTCTACATAGCCATCGC
STS size: 150

```



<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L12085  
Center clone name: 179\_J\_16  
-----

\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 737: contig of 737 bp in length  
\* 738 837: gap of 100 bp  
\* 838 1551: contig of 714 bp in length  
\* 1552 1651: gap of 100 bp  
\* 1652 2343: contig of 692 bp in length  
\* 2344 2443: gap of 100 bp  
\* 2444 3138: contig of 695 bp in length  
\* 3139 3238: gap of 100 bp  
\* 3239 3953: contig of 715 bp in length  
\* 3954 4053: gap of 100 bp  
\* 4054 4722: contig of 669 bp in length  
\* 4723 4822: gap of 100 bp  
\* 4823 5530: contig of 708 bp in length  
\* 5531 5630: gap of 100 bp  
\* 5631 6330: contig of 700 bp in length  
\* 6331 6430: gap of 100 bp  
\* 6431 7165: contig of 735 bp in length  
\* 7166 7265: gap of 100 bp  
\* 7266 7957: contig of 692 bp in length  
\* 7958 8057: gap of 100 bp  
\* 8058 8788: contig of 731 bp in length  
\* 8789 8888: gap of 100 bp  
\* 8889 9619: contig of 731 bp in length  
\* 9620 9719: gap of 100 bp  
\* 9720 10467: contig of 748 bp in length  
\* 10468 10567: gap of 100 bp  
\* 10568 11265: contig of 698 bp in length  
\* 11266 11365: gap of 100 bp  
\* 11366 12055: contig of 690 bp in length  
\* 12056 12155: gap of 100 bp  
\* 12156 12847: contig of 692 bp in length  
\* 12848 12947: gap of 100 bp  
\* 12948 13662: contig of 715 bp in length  
\* 13663 13762: gap of 100 bp  
\* 13763 14476: contig of 714 bp in length  
\* 14477 14576: gap of 100 bp  
\* 14577 15302: contig of 726 bp in length  
\* 15303 15402: gap of 100 bp  
\* 15403 16105: contig of 703 bp in length  
\* 16106 16205: gap of 100 bp  
\* 16206 16942: contig of 737 bp in length  
\* 16943 17042: gap of 100 bp  
\* 17043 17786: contig of 744 bp in length  
\* 17787 17886: gap of 100 bp  
\* 17887 18612: contig of 726 bp in length  
\* 18613 18712: gap of 100 bp  
\* 18713 19420: contig of 708 bp in length  
\* 19421 19520: gap of 100 bp  
\* 19521 20216: contig of 696 bp in length  
\* 20217 20316: gap of 100 bp  
\* 20317 21022: contig of 706 bp in length  
\* 21023 21122: gap of 100 bp

\* 21123 21827: contig of 705 bp in length  
\* 21828 21927: gap of 100 bp  
\* 21928 22639: contig of 712 bp in length  
\* 22640 22739: gap of 100 bp  
\* 22740 23453: contig of 714 bp in length  
\* 23454 23553: gap of 100 bp  
\* 23554 24279: contig of 726 bp in length  
\* 24280 24379: gap of 100 bp  
\* 24380 25082: contig of 703 bp in length  
\* 25083 25182: gap of 100 bp  
\* 25183 25907: contig of 725 bp in length  
\* 25908 26007: gap of 100 bp  
\* 26008 26732: contig of 725 bp in length  
\* 26733 26832: gap of 100 bp  
\* 26833 27561: contig of 729 bp in length  
\* 27562 27661: gap of 100 bp  
\* 27662 28358: contig of 697 bp in length  
\* 28359 28458: gap of 100 bp  
\* 28459 29136: contig of 678 bp in length  
\* 29137 29236: gap of 100 bp  
\* 29237 29936: contig of 700 bp in length  
\* 29937 30036: gap of 100 bp  
\* 30037 30736: contig of 700 bp in length  
\* 30737 30836: gap of 100 bp  
\* 30837 31561: contig of 725 bp in length  
\* 31562 31661: gap of 100 bp  
\* 31662 32380: contig of 719 bp in length  
\* 32381 32480: gap of 100 bp  
\* 32481 33224: contig of 744 bp in length  
\* 33225 33324: gap of 100 bp  
\* 33325 34053: contig of 729 bp in length  
\* 34054 34153: gap of 100 bp  
\* 34154 34877: contig of 724 bp in length  
\* 34878 34977: gap of 100 bp  
\* 34978 35703: contig of 726 bp in length  
\* 35704 35803: gap of 100 bp  
\* 35804 36503: contig of 700 bp in length  
\* 36504 36603: gap of 100 bp  
\* 36604 37315: contig of 712 bp in length  
\* 37316 37415: gap of 100 bp  
\* 37416 38120: contig of 705 bp in length  
\* 38121 38220: gap of 100 bp  
\* 38221 38916: contig of 696 bp in length  
\* 38917 39016: gap of 100 bp  
\* 39017 39712: contig of 696 bp in length  
\* 39713 39812: gap of 100 bp  
\* 39813 40535: contig of 723 bp in length  
\* 40536 40635: gap of 100 bp  
\* 40636 41335: contig of 700 bp in length  
\* 41336 41435: gap of 100 bp  
\* 41436 42087: contig of 652 bp in length  
\* 42088 42187: gap of 100 bp  
\* 42188 42914: contig of 727 bp in length  
\* 42915 43014: gap of 100 bp  
\* 43015 43738: contig of 724 bp in length  
\* 43739 43838: gap of 100 bp  
\* 43839 44553: contig of 715 bp in length  
\* 44554 44653: gap of 100 bp  
\* 44654 45353: contig of 700 bp in length  
\* 45354 45453: gap of 100 bp  
\* 45454 46147: contig of 694 bp in length  
\* 46148 46247: gap of 100 bp  
\* 46248 46939: contig of 692 bp in length  
\* 46940 47039: gap of 100 bp  
\* 47040 47730: contig of 691 bp in length  
\* 47731 47830: gap of 100 bp  
\* 47831 48537: contig of 707 bp in length  
\* 48538 48637: gap of 100 bp  
\* 48638 49354: contig of 717 bp in length  
\* 49355 49454: gap of 100 bp  
\* 49455 50194: contig of 740 bp in length  
\* 50195 50294: gap of 100 bp  
\* 50295 51026: contig of 732 bp in length



\* 20217 20316: gap of 100 bp in length  
 \* 20317 21022: contig of 706 bp in length  
 \* 21023 21122: gap of 100 bp in length  
 \* 21123 21827: contig of 705 bp in length  
 \* 21828 21927: gap of 100 bp in length  
 \* 21928 22639: contig of 712 bp in length  
 \* 22640 22739: gap of 100 bp in length  
 \* 22740 23453: contig of 714 bp in length  
 \* 23454 23553: gap of 100 bp in length  
 \* 23554 24279: contig of 726 bp in length  
 \* 24280 24379: gap of 100 bp in length  
 \* 24380 25082: contig of 703 bp in length  
 \* 25083 25182: gap of 100 bp in length  
 \* 25183 25907: contig of 725 bp in length  
 \* 25908 26007: gap of 100 bp in length  
 \* 26008 26732: contig of 725 bp in length  
 \* 26733 26832: gap of 100 bp in length  
 \* 26833 27561: contig of 729 bp in length  
 \* 27562 27661: gap of 100 bp in length  
 \* 27662 28358: contig of 697 bp in length  
 \* 28359 28458: gap of 100 bp in length  
 \* 28459 29136: contig of 678 bp in length  
 \* 29137 29236: gap of 100 bp in length  
 \* 29237 29936: contig of 700 bp in length  
 \* 29937 30036: gap of 100 bp in length  
 \* 30037 30736: contig of 700 bp in length  
 \* 30737 30836: contig of 725 bp in length  
 \* 30837 31561: contig of 725 bp in length  
 \* 31562 31661: gap of 100 bp in length  
 \* 32381 32480: contig of 719 bp in length  
 \* 32481 33224: contig of 744 bp in length  
 \* 33225 33324: gap of 100 bp in length  
 \* 33225 34053: contig of 729 bp in length  
 \* 34054 34153: gap of 100 bp in length  
 \* 34154 34877: contig of 724 bp in length  
 \* 34878 34977: gap of 100 bp in length  
 \* 34978 35703: contig of 726 bp in length  
 \* 35704 35803: gap of 100 bp in length  
 \* 35804 36503: contig of 700 bp in length  
 \* 36504 36603: gap of 100 bp in length  
 \* 36604 37315: contig of 712 bp in length  
 \* 37316 37415: gap of 100 bp in length  
 \* 37416 38120: contig of 705 bp in length  
 \* 38121 38220: gap of 100 bp in length  
 \* 38221 38916: contig of 696 bp in length  
 \* 38917 39016: gap of 100 bp in length  
 \* 39017 39712: contig of 696 bp in length  
 \* 39713 39812: gap of 100 bp in length  
 \* 39813 40535: contig of 723 bp in length  
 \* 40536 40635: gap of 100 bp in length  
 \* 40636 41335: contig of 700 bp in length  
 \* 41336 41435: gap of 100 bp in length  
 \* 41436 42087: contig of 652 bp in length  
 \* 42088 42187: gap of 100 bp in length  
 \* 42188 42914: contig of 727 bp in length  
 \* 42915 43014: gap of 100 bp in length  
 \* 43015 43738: contig of 724 bp in length  
 \* 43739 43838: gap of 100 bp in length  
 \* 43839 44533: contig of 715 bp in length  
 \* 44534 44633: gap of 100 bp in length  
 \* 44634 45353: contig of 700 bp in length  
 \* 45354 45453: gap of 100 bp in length  
 \* 45454 46147: contig of 694 bp in length  
 \* 46148 46247: gap of 100 bp in length  
 \* 46248 46939: contig of 692 bp in length  
 \* 46940 47039: gap of 100 bp in length  
 \* 47040 47730: contig of 691 bp in length  
 \* 47731 47830: gap of 100 bp in length  
 \* 47831 48537: contig of 707 bp in length  
 \* 48538 48637: gap of 100 bp in length  
 \* 48638 49354: contig of 717 bp in length  
 \* 49355 49454: gap of 100 bp in length

\* 49455 50194: contig of 740 bp in length  
 \* 50195 50294: gap of 100 bp in length  
 \* 50295 51026: contig of 732 bp in length  
 \* 51027 51126: gap of 100 bp in length  
 \* 51127 51836: contig of 710 bp in length  
 \* 51837 51936: gap of 100 bp in length  
 \* 51937 52671: contig of 735 bp in length  
 \* 52672 52771: gap of 100 bp in length  
 \* 52772 53502: contig of 731 bp in length  
 \* 53503 53602: gap of 100 bp in length  
 \* 53603 54343: contig of 741 bp in length  
 \* 54344 54443: gap of 100 bp in length  
 \* 54444 55141: contig of 698 bp in length  
 \* 55142 55241: gap of 100 bp in length  
 \* 55242 55950: contig of 709 bp in length  
 \* 55951 56050: gap of 100 bp in length

Query Match 2.3%; Score 120; DB 77; Length 69740;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-53;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 agagcaatgataatgagcgagcgctcacaagtgtttaaactactggcaaaatgtttgggg 969  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 28043 AGAGCAATGATAATGAGGAGCGCCTACAAAGTTGTAAACTACTGGCAAAATGTTGGG 27984  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 970 caaagattcagaattgcttctcaaaacagccacttgcagctgacttggcaggt 1029  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 27983 CAAAGGATTCAGAATTGGCTTCTCAAAACAGCCACTTTGGCAGTGTACTTTGGCAGGT 27924

## RESULT 16

AC087748 69740 bp DNA HTG 22-JAN-2001  
 Homo sapiens chromosome 18 clone CTFB-179J16 map 18, LOW-PASS  
 SEQUENCE SAMPLING.

AC087748  
 VERSION AC087748.1 GI:12331536  
 KEYWORDS HTG: HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 69740)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 18, clone CTFB-179J16  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 69740)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galegan,J.,  
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,  
 Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,  
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheeters,R., Meldrum,J., Meneus,L., Mihova,I., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
 Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo.A.,  
 Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G., Zainoun,J.,  
 Zemdek,L., Zimmer,A. and Zody,M.  
 Direct Submission

TITLE  
 JOURNAL  
 COMMENT Submitted (22-JAN-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)



```

/note="assembly_fragment:00874
fragment_chain:2"
misc_feature 147136..149710
/note="assembly_fragment:00391
fragment_chain:3"
misc_feature 149811..163642
/note="assembly_fragment:00398
fragment_chain:3"
BASE COUNT 54756 a 31368 c 29374 g 47337 t 807 others
ORIGIN

Query Match 3.0%; Score 156; DB 79; Length 163642;
Best Local Similarity 99.5%; Pred. No. 1.8e-72;
Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 172 agatgtgtgaaactttatgatatgacacgagactctgaagaaagagacgttt 231
|||||
Db 32938 AGATGTTGTGAACACTTTATGGATATGGACGAGACTCTGAAGAAGAAAGAGCCTTT 32879
|||||

QY 232 attaaacactagctttacatctgtctcagattttttctcaagcatctcgtgtaagatg 291
|||||
Db 32878 ATTAAACCTAGCTTTACATCTTGCCTTCAGATTTTCTCAAGCATCTCGATAAAGATG 32819
|||||

QY 292 ttgcgttaactgtagctcgtcctgtcgatatttttcaggattttatgctcctgaagctc 351
|||||
Db 32818 TTGCGTTACTGGTAGCTGCTGCTTGCATATATTTTCAGGATTTATGCTCTCGAAGCTC 32759
|||||

QY 352 cttacacatccctcgtataaactaaagg 378
|||||
Db 32758 CTTACACATCCCTGATAACTAAAGG 32732
|||||

RESULT 15
AC087748/c
LOCUS AC087748 69740 bp DNA HTG 22-JAN-2001
DEFINITION Homo sapiens chromosome 18 clone CTB-179J16 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087748
VERSION AC087748.1 GI:12331536
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69740)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone CTB-179J16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 69740)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
JOURNAL Submitted (22-JAN-2001) Whitehead Institute/MIT Center for Genome

```

## COMMENT

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12085
Center clone name: 179_J16
-----
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 737: contig of 737 bp in length
* 738 837: gap of 100 bp
* 838 1551: contig of 714 bp in length
* 1552 1651: gap of 100 bp
* 1652 2343: contig of 692 bp in length
* 2344 2443: gap of 100 bp
* 2444 3138: contig of 695 bp in length
* 3139 3238: gap of 100 bp
* 3239 3953: contig of 715 bp in length
* 3954 4053: gap of 100 bp
* 4054 4722: contig of 669 bp in length
* 4723 4822: gap of 100 bp
* 4823 5530: contig of 708 bp in length
* 5531 5630: gap of 100 bp
* 5631 6330: contig of 700 bp in length
* 6331 6430: gap of 100 bp
* 6431 7165: contig of 735 bp in length
* 7166 7265: gap of 100 bp
* 7266 7957: contig of 692 bp in length
* 7958 8057: gap of 100 bp
* 8058 8788: contig of 731 bp in length
* 8789 8888: gap of 100 bp
* 8889 9619: contig of 731 bp in length
* 9620 9719: gap of 100 bp
* 9720 10467: contig of 748 bp in length
* 10468 10567: gap of 100 bp
* 10568 11265: contig of 698 bp in length
* 11266 11365: gap of 100 bp
* 11366 12055: contig of 690 bp in length
* 12056 12155: gap of 100 bp
* 12156 12847: contig of 692 bp in length
* 12848 12947: gap of 100 bp
* 12948 13662: contig of 715 bp in length
* 13663 13762: gap of 100 bp
* 13763 14476: contig of 714 bp in length
* 14477 14576: gap of 100 bp
* 14577 15302: contig of 726 bp in length
* 15303 15402: gap of 100 bp
* 15403 16105: contig of 703 bp in length
* 16106 16205: gap of 100 bp
* 16206 16942: contig of 737 bp in length
* 16943 17042: gap of 100 bp
* 17043 17786: contig of 744 bp in length
* 17787 17886: gap of 100 bp
* 17887 18612: contig of 726 bp in length
* 18613 18712: gap of 100 bp
* 18713 19420: contig of 708 bp in length
* 19421 19520: gap of 100 bp
* 19521 20216: contig of 696 bp in length

```











```

QY 4255 cattcaaaccaatttcaaaattatttgcaaaagtcttcaaaatttgtaaacatacatatt 4314
|||||
DB 34062 CATTTCACCAAAATTTCAAAATTTATTTGCAAAAGTTCTTAAATTTGTAACATACATATT 34121
|||||

QY 4315 gctgtattaaatccatattatttagccattacacatagtagta 4357
|||||
DB 34122 GCTGTATTAAATTCATATATTTAGCCCATACACTAGTAGTA 34164
|||||

RESULT 12
HS267P19
LOCUS
DEFINITION
  HS267P19 113704 bp DNA 22-NOV-1999
  Human DNA sequence from cosmid 267P19, BRCA2 gene region chromosome
  13q12-13 contains polymorphic CA repeat.
ACCESSION
  275889
VERSION
  275889.1 GI:1430780
KEYWORDS
  13q12-13; repeat polymorphism.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 113704)
  Odeh, C.
  Direct Submission
  Submitted (04-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1RO, UK. E-mail enquires: humquery@sanger.ac.uk
  requests: clonesrequest@sanger.ac.uk
  IMPORTANT:
  This sequence is not the entire insert of clone 267P19. It may be
  shorter because we only sequence overlapping sections once, or
  longer because we arrange for a small overlap between neighbouring
  submissions.
  This sequence has been finished according to sequence map criteria
  as follows. An attempt is made to resolve all sequencing problems,
  such as compressions and repeats, but not necessarily within known
  annotated human repeat sequence elements (e.g. Alu). Where the
  sequence is ambiguous, there is an annotation using the 'unsure'
  feature key.
  The true left end of clone 267P19 is at 1 in this sequence. The
  true right end of clone 26H23 is at 24539.
  The true left end of clone 49J10 is at 113606.
  267P19 is from the human PAC library.
FEATURES
  source
    1..113704
    /organism="Homo sapiens"
    /db_xref="taxon.9606"
    /chromosome="13"
    /map="13q12-13"
    /clone="XX-267P19"
  repeat_region
    1..235
    /partial
    /note="Alu repeat: matches 252. .1 of consensus"
  repeat_region
    370..662
    /partial
    /note="Alu repeat: matches 308. .1 of consensus"
  repeat_region
    4849..5140
    /partial
    /note="Alu repeat: matches 308. .1 of consensus"
  repeat_region
    5367..5660
    /note="Alu repeat: matches 308. .1 of consensus"
  repeat_region
    6080..6187
    /note="Alu repeat: matches 1. .308 of consensus"
  repeat_region
    8133..8401
    /note="2 copies of 54 mer 87 & conserved"
  repeat_region
    /note="MER7A element fragment"
  repeat_region
    8623..8748
    /note="L1 element fragment"
  repeat_region
    9617..9711
    /partial
    /note="Alu repeat: matches 305. .210 of consensus"
  repeat_region
    9724..9876
    /partial
    /note="Alu repeat: matches 150. .1 of consensus"
  repeat_region
    10567..10594
    /note="Alu repeat: matches 1. .308 of consensus"

```

```

repeat_region
  /note="14 copies of 2 mer 93 & conserved"
  10616..10900
  /partial
  /note="Alu repeat: matches 300. .1 of consensus"
  10960..10985
  /note="13 copies of 2 mer 92 & conserved"
  11912..12100
  /partial
  /note="Alu repeat: matches 268. .65 of consensus"
  17050..17359
  /note="Alu repeat: matches 1. .306 of consensus"
  17639..17888
  /partial
  /note="Alu repeat: matches 34. .306 of consensus"
  19151..19980
  /note="L1 element fragment"
  19981..20264
  /partial
  /note="Alu repeat: matches 308. .1 of consensus"
  20298..21190
  /note="L1 element fragment"
  21199..21488
  /partial
  /note="Alu repeat: matches 308. .1 of consensus"
  21538..22226
  /note="L1 element fragment"
  25696..25989
  /note="Alu repeat: matches 1. .308 of consensus"
  26902..27167
  /note="2 copies of 133 mer 93 & conserved"
  27850..27930
  /note="L1 element fragment"
  28002..28043
  /note="L1 element fragment"
  28323..28780
  /note="L1 element fragment"
  29176..29449
  /note="Alu repeat: matches 1. .306 of consensus"
  29904..30192
  /note="Alu repeat: matches 1. .308 of consensus"
  30354..30663
  /note="L1 element fragment"
  30683..30972
  /partial
  /note="Alu repeat: matches 308. .1 of consensus"
  30977..31204
  /note="L1 element fragment"
  34819..35121
  /partial
  /note="Alu repeat: matches 308. .1 of consensus"
  37059..37282
  /note="MUT2B2 element fragment"
  37453..37506
  /note="MUT2B2 element fragment"
  38414..38706
  /partial
  /note="Alu repeat: matches 308. .1 of consensus"
  39516..39563
  /note="24 copies of 2 mer 81 & conserved"
  41244..41535
  /note="Alu repeat: matches 1. .308 of consensus"
  43994..44242
  /partial
  /note="Alu repeat: matches 44. .308 of consensus"
  complement(44369..44677)
  /note="match: 217151 DNA segment containing (CA) repeat"
  complement(44479..44510)
  /note="16 copies of CA 100% conserved; Polymorphic by
  comparison with 217151"
  44702..44889
  /note="2 copies of 94 mer 85 & conserved"
  45935..46229
  /note="Alu repeat: matches 1. .308 of consensus"

```











## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campolano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGuck,K., McKernan,K., McPheeters,R.,  
 Meldrin,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Ye,W.J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L10161  
 Center clone name: 660\_R\_19

\* NOTE: This record contains 68 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 685: contig of 685 bp in length  
 \* 686 785: gap of 100 bp  
 \* 786 1513: contig of 728 bp in length  
 \* 1514 1613: gap of 100 bp  
 \* 1614 2354: contig of 741 bp in length  
 \* 2355 2454: gap of 100 bp  
 \* 2455 3184: contig of 730 bp in length  
 \* 3185 3284: gap of 100 bp  
 \* 3285 3959: contig of 675 bp in length  
 \* 3960 4059: gap of 100 bp  
 \* 4060 4746: contig of 687 bp in length  
 \* 4747 4846: gap of 100 bp  
 \* 4847 5564: contig of 718 bp in length  
 \* 5565 5664: gap of 100 bp  
 \* 5665 6360: contig of 696 bp in length  
 \* 6361 6460: gap of 100 bp  
 \* 6461 7132: contig of 672 bp in length  
 \* 7133 7232: gap of 100 bp  
 \* 7233 7937: contig of 705 bp in length  
 \* 7938 8037: gap of 100 bp  
 \* 8038 8749: contig of 712 bp in length  
 \* 8750 8849: gap of 100 bp  
 \* 8850 9522: contig of 673 bp in length  
 \* 9523 9622: gap of 100 bp  
 \* 9623 10356: contig of 734 bp in length  
 \* 10357 10456: gap of 100 bp

10457 11181: contig of 725 bp in length  
 \* 11182 11281: gap of 100 bp  
 \* 11282 11975: contig of 694 bp in length  
 \* 11976 12075: gap of 100 bp  
 \* 12076 12785: contig of 710 bp in length  
 \* 12786 12885: gap of 100 bp  
 \* 12886 13557: contig of 672 bp in length  
 \* 13558 13657: gap of 100 bp  
 \* 13658 14347: contig of 690 bp in length  
 \* 14348 14447: gap of 100 bp  
 \* 14448 15132: contig of 685 bp in length  
 \* 15133 15232: gap of 100 bp  
 \* 15233 15912: contig of 680 bp in length  
 \* 15913 16012: gap of 100 bp  
 \* 16013 16731: contig of 719 bp in length  
 \* 16732 16831: gap of 100 bp  
 \* 16832 17506: contig of 675 bp in length  
 \* 17507 17606: gap of 100 bp  
 \* 17607 18339: contig of 733 bp in length  
 \* 18340 18439: gap of 100 bp  
 \* 18440 19109: contig of 670 bp in length  
 \* 19110 19209: gap of 100 bp  
 \* 19210 19900: contig of 691 bp in length  
 \* 19901 20000: gap of 100 bp  
 \* 20001 20693: contig of 693 bp in length  
 \* 20694 20793: gap of 100 bp  
 \* 20794 21464: contig of 671 bp in length  
 \* 21465 21564: gap of 100 bp  
 \* 21565 22245: contig of 681 bp in length  
 \* 22246 22345: gap of 100 bp  
 \* 22346 23049: contig of 704 bp in length  
 \* 23050 23149: gap of 100 bp  
 \* 23150 23857: contig of 708 bp in length  
 \* 23858 23957: gap of 100 bp  
 \* 23958 24685: contig of 728 bp in length  
 \* 24686 24785: gap of 100 bp  
 \* 24786 25499: contig of 714 bp in length  
 \* 25500 25599: gap of 100 bp  
 \* 25600 26302: contig of 703 bp in length  
 \* 26303 26402: gap of 100 bp  
 \* 26403 27108: contig of 706 bp in length  
 \* 27109 27208: gap of 100 bp  
 \* 27209 27910: contig of 702 bp in length  
 \* 27911 28010: gap of 100 bp  
 \* 28011 28720: contig of 710 bp in length  
 \* 28721 28820: gap of 100 bp  
 \* 28821 29484: contig of 664 bp in length  
 \* 29485 29584: gap of 100 bp  
 \* 29585 30328: contig of 744 bp in length  
 \* 30329 30428: gap of 100 bp  
 \* 30429 31111: contig of 683 bp in length  
 \* 31112 31211: gap of 100 bp  
 \* 31212 31931: contig of 720 bp in length  
 \* 31932 32031: gap of 100 bp  
 \* 32032 32710: contig of 679 bp in length  
 \* 32711 32810: gap of 100 bp  
 \* 32811 33502: contig of 692 bp in length  
 \* 33503 33602: gap of 100 bp  
 \* 33603 34301: contig of 699 bp in length  
 \* 34302 34401: gap of 100 bp  
 \* 34402 35100: contig of 699 bp in length  
 \* 35101 35200: gap of 100 bp  
 \* 35201 35914: contig of 714 bp in length  
 \* 35915 36014: gap of 100 bp  
 \* 36015 36709: contig of 695 bp in length  
 \* 36710 36809: gap of 100 bp  
 \* 36810 37518: contig of 709 bp in length  
 \* 37519 37618: gap of 100 bp  
 \* 37619 38330: contig of 712 bp in length  
 \* 38331 38430: gap of 100 bp  
 \* 38431 39109: contig of 679 bp in length  
 \* 39110 39209: gap of 100 bp  
 \* 39210 39912: contig of 703 bp in length



\* 15913 16012: gap of 100 bp  
 \* 16013 16731: contig of 719 bp in length  
 \* 16732 16831: gap of 100 bp  
 \* 16832 17506: contig of 675 bp in length  
 \* 17507 17606: gap of 100 bp  
 \* 17607 18339: contig of 733 bp in length  
 \* 18340 18439: gap of 100 bp  
 \* 18440 19109: contig of 670 bp in length  
 \* 19110 19209: gap of 100 bp  
 \* 19210 19900: contig of 691 bp in length  
 \* 19901 20000: gap of 100 bp  
 \* 20001 20693: contig of 693 bp in length  
 \* 20694 20793: gap of 100 bp  
 \* 20794 21464: contig of 671 bp in length  
 \* 21465 21564: gap of 100 bp  
 \* 21565 22245: contig of 681 bp in length  
 \* 22246 22345: gap of 100 bp  
 \* 22346 23049: contig of 704 bp in length  
 \* 23050 23149: gap of 100 bp  
 \* 23150 23857: contig of 708 bp in length  
 \* 23858 23957: gap of 100 bp  
 \* 23958 24685: contig of 728 bp in length  
 \* 24686 24785: gap of 100 bp  
 \* 24786 25499: contig of 714 bp in length  
 \* 25500 25599: gap of 100 bp  
 \* 25600 26302: contig of 703 bp in length  
 \* 26303 26402: gap of 100 bp  
 \* 26403 27108: contig of 706 bp in length  
 \* 27109 27208: gap of 100 bp  
 \* 27209 27910: contig of 702 bp in length  
 \* 27911 28010: gap of 100 bp  
 \* 28011 28720: contig of 710 bp in length  
 \* 28721 28820: gap of 100 bp  
 \* 28821 29484: contig of 664 bp in length  
 \* 29485 29584: gap of 100 bp  
 \* 29585 30328: contig of 744 bp in length  
 \* 30329 30428: gap of 100 bp  
 \* 30429 31111: contig of 683 bp in length  
 \* 31112 31211: gap of 100 bp  
 \* 31212 31931: contig of 720 bp in length  
 \* 31932 32031: gap of 100 bp  
 \* 32032 32710: contig of 679 bp in length  
 \* 32711 32810: gap of 100 bp  
 \* 32811 33502: contig of 692 bp in length  
 \* 33503 33602: gap of 100 bp  
 \* 33603 34301: contig of 699 bp in length  
 \* 34302 34401: gap of 100 bp  
 \* 34402 35100: contig of 699 bp in length  
 \* 35101 35200: gap of 100 bp  
 \* 35201 35914: contig of 714 bp in length  
 \* 35915 36014: gap of 100 bp  
 \* 36015 36709: contig of 695 bp in length  
 \* 36710 36809: gap of 100 bp  
 \* 36810 37518: contig of 709 bp in length  
 \* 37519 37618: gap of 100 bp  
 \* 37619 38330: contig of 712 bp in length  
 \* 38331 38430: gap of 100 bp  
 \* 38431 39109: contig of 679 bp in length  
 \* 39110 39209: gap of 100 bp  
 \* 39210 39912: contig of 703 bp in length  
 \* 39913 40012: gap of 100 bp  
 \* 40013 40712: contig of 700 bp in length  
 \* 40713 40812: gap of 100 bp  
 \* 40813 41514: contig of 702 bp in length  
 \* 41515 41614: gap of 100 bp  
 \* 41615 42294: contig of 680 bp in length  
 \* 42295 42394: gap of 100 bp  
 \* 42395 43094: contig of 700 bp in length  
 \* 43095 43194: gap of 100 bp  
 \* 43195 43915: contig of 721 bp in length  
 \* 43916 44015: gap of 100 bp  
 \* 44016 44740: contig of 725 bp in length  
 \* 44741 44840: gap of 100 bp

\* 44841 45533: contig of 693 bp in length  
 \* 45534 45633: gap of 100 bp  
 \* 45634 46330: contig of 697 bp in length  
 \* 46331 46430: gap of 100 bp  
 \* 46431 47137: contig of 707 bp in length  
 \* 47138 47237: gap of 100 bp  
 \* 47238 47927: contig of 690 bp in length  
 \* 47928 48027: gap of 100 bp  
 \* 48028 48709: contig of 682 bp in length  
 \* 48710 48809: gap of 100 bp  
 \* 48810 49509: contig of 700 bp in length  
 \* 49510 49609: gap of 100 bp  
 \* 49610 50331: contig of 722 bp in length  
 \* 50332 50431: gap of 100 bp  
 \* 50432 51135: contig of 704 bp in length  
 \* 51136 51235: gap of 100 bp  
 \* 51236 51981: contig of 746 bp in length  
 \* 51982 52081: gap of 100 bp  
 \* 52082 52783: contig of 702 bp in length  
 \* 52784 52883: gap of 100 bp  
 \* 52884 53580: contig of 697 bp in length  
 \* 53581 53680: gap of 100 bp  
 \* 53681 54398: contig of 718 bp in length.

# FEATURES

Source  
 1. 54398  
 /organism="Homo sapiens"

Query Match 5.9%; Score 311; DB 73; Length 54398;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-156;  
 Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3819 agagggccatcgctcagaaatctgatgaacagcagtggtgctgaggaagaaagagctcaaa 3878  
 |||||

Db 29407 AGAGGCCATACGGCTTCAGAAATCTGTGAACAGCAGTGGCTGTGAGGAAAGAGCTCAA 29348  
 |||||

QY 3879 gaagatatattagaaatgaagatgaacagaatagtcgcgcaaaaaagggtaaaagaggc 3938  
 |||||

Db 29347 GAAGATATATTAGAAATGAAGATGAACAGANTAGTCCGCCAAAAGGGTAAAGAGGC 29288  
 |||||

QY 3939 cgaccaccaaaaccttggtaggtacaccacaaagaagccacaaatgaaactct 3998  
 |||||

Db 29287 CGACCACCAAAAGCTCTTGGTGGAGGTACACCAAAAGAGCACCACCAATGAAACTTCT 29228  
 |||||

QY 3999 aaaaaaggaagcaaaaaaatactgacctccagcaccagagaggaagaagaaga 4058  
 |||||

Db 29227 AAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCAGGAGGAGGAAGAAGAA 29168  
 |||||

QY 4059 agacaaagtgaatacggacagaaagtcacaaagcaaacagaccagtgctcaagaga 4118  
 |||||

Db 29167 AGACAAAGTGAATACGGAACAGAGAGTCCAAAAGCAACAGCACCAGTGCTCAAGGAGA 29108  
 |||||

QY 4119 gcacagcagag 4129  
 |||||

Db 29107 GCACAGCAGAG 29097

## RESULT 10

AC068224

LOCUS

DEFINITION

Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS

SEQUENCE SAMPLING.

AC068224

VERSION

AC068224.1 GI:7671284

KEYWORDS

HTG; HTGS\_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 54398)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Unpublished

JOURNAL

REFERENCE

2 (bases 1 to 54398)



```

* 60893 61663: contig of 771 bp in length
* 61664 61763: gap of 100 bp
* 61764 62539: contig of 776 bp in length
* 62540 62639: gap of 100 bp
* 62640 63422: contig of 783 bp in length

Query Match      8.4%; Score 441; DB 64; Length 72157;
Best Local Similarity 100.0%; Pred. No. 1.4e-226;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3689 gtctaattgagaagcctagaggcagagaaaaacgcccgtcacagaaacagagagagaa 3748
|||||
Dp 39686 GTCCTAATGGAAAGCCCTAGAGGAGGAGAAAACGCCCGTCACAGACAGAGGAGAA 39627
|||||
QY 3749 attagtgatggatgactgaacttggtacagaaacacctaagaggcagtcagcg 3808
|||||
Dp 39626 ATTAGGTATGGATGACTTGACCTAAGCTGTGTACAGAAACAGAAACCTAAAGGCAGTCAGCG 39567
|||||
QY 3809 aagtcggaagagccatagcgttcagatctctatgaacgagcgtgacctgagagaaaa 3868
|||||
Dp 39566 AAGTCGGAAGAGGCCATACGCGCTTCAGAACTCTGATGAACACGACGATGCGCTTGAGGAAAA 39507
|||||
QY 3869 gaggctcaaaagatatattagaaaatgaagataacagaaatagtcgcccaaaaaaggg 3928
|||||
Dp 39506 GAGGCTCAAGAGATATATTAGAAATGAAGATGAACAGATGTCCCCAAAAGGG 39447
|||||
QY 3929 taaagagcgccaccacaaacaccttggtaggtacacccaagaagagcccaacaat 3988
|||||
Dp 39446 TAAAGAGCGCCACCACCAAAACCTCTTGGTGGAGGTACACCAAAAAGAGAGCCCAACAT 39387
|||||
QY 3989 gaaactcttaaaagaagcaaaaaaaatcttgacctccagcaccagagagagga 4048
|||||
Dp 39386 GAAAACTTCTTAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGAGAGGAGGA 39327
|||||
QY 4049 aqaagaagaagcaaaagtggaaatacaggaacagagtgccaaagcaaacagaccgagt 4108
|||||
Dp 39326 AGAAGAGAAAGACAAAGTGGAAATACGGACAGAAAGTCCAAAGCAACACACCCGAGT 39267
|||||
QY 4109 gtcaaggagagcacagcagag 4129
|||||
Dp 39266 GTCAAGGAGACACAGCAGAG 39246
|||||

RESULT 9
AC068224/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC068224
VERSION AC068224.1 GI:7671284
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 54398)
BIRREN, B., LINTON, L., NUSBAUM, C., LANDER, E., ABRAHAM, H., ALLEN, N.,
ANDERSON, S., BALDWIN, J., BARNA, N., BASTIEN, V., BEDA, F.,
BOGUSLAVSKIY, L., BOUKHAGALTER, B., BROWN, A., BURKETT, G.,
CAMPOLANO, A., CASTLE, A., CHOEPPEL, Y., COLANGELO, M., COLLINS, S.,
COLLYMORE, A., COOKE, P., DEARELLANO, K., DEWAR, K., DIAZ, J. S.,
DODGE, S., DOMINO, M., DOYLE, M., FERREIRA, P., FITZHUGH, W., GAGE, D.,
GALAGAN, J., GARDYNA, S., GINDE, S., GOYETTE, M., GRAHAM, L.,
GRAND-PIERRE, N., GRANT, G., HAGOS, B., HEAFORD, A., HORTON, L.,
HOWLAND, J. C., ILIEV, I., JOHNSON, R., JONES, C., KANN, L., KARATAS, A.,
KLEIN, J., LAKOCQUE, K., LANAZARES, R., LANDERS, T., LEHOCZKY, J.,
LEVINE, R., LIEU, C., LIU, G., LOCKE, K., MACDONALD, P., MARQUIS, N.,
MCARTHUR, M., MCEWAN, P., MCGURK, A., MCKERNAN, K., MCPHEETERS, R.,
MELDRIM, J., MENEUS, L., MIHOVA, T., MIRANDA, C., MLENGA, V., MORROW, J.,

```

```

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L10161
Center clone name: 660_H_19
-----
* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 685: contig of 685 bp in length
* 686 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp
* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 12885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp
* 15233 15912: contig of 680 bp in length

```

TITLE  
JOURNAL

COMMENT



\* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 757: contig of 757 bp in length  
 758 857: gap of 100 bp  
 858 1635: contig of 778 bp in length  
 1636 1735: gap of 100 bp  
 1736 2490: contig of 755 bp in length  
 2491 2590: gap of 100 bp  
 2591 3343: contig of 753 bp in length  
 3344 3443: gap of 100 bp  
 3444 4224: contig of 781 bp in length  
 4225 4324: gap of 100 bp  
 4325 5088: contig of 764 bp in length  
 5089 5188: gap of 100 bp  
 5189 5968: contig of 780 bp in length  
 5969 6068: gap of 100 bp  
 6069 6844: contig of 776 bp in length  
 6845 6944: gap of 100 bp  
 6945 7701: contig of 757 bp in length  
 7702 7801: gap of 100 bp  
 7802 8573: contig of 772 bp in length  
 8574 8673: gap of 100 bp  
 8674 9430: contig of 757 bp in length  
 9431 9530: gap of 100 bp  
 9531 10295: contig of 765 bp in length  
 10296 10395: gap of 100 bp  
 10396 11139: contig of 744 bp in length  
 11140 11239: gap of 100 bp  
 11240 11988: contig of 749 bp in length  
 11989 12088: gap of 100 bp  
 12089 12874: contig of 786 bp in length  
 12875 12974: gap of 100 bp  
 12975 13750: contig of 776 bp in length  
 13751 13850: gap of 100 bp  
 13851 14638: contig of 788 bp in length  
 14639 14738: gap of 100 bp  
 14739 15520: contig of 782 bp in length  
 15521 15620: gap of 100 bp  
 15621 16381: contig of 761 bp in length  
 16382 16481: gap of 100 bp  
 16482 17248: contig of 767 bp in length  
 17249 17348: gap of 100 bp  
 17349 18123: contig of 775 bp in length  
 18124 18223: gap of 100 bp  
 18224 18977: contig of 754 bp in length  
 18978 19077: gap of 100 bp  
 19078 19860: contig of 783 bp in length  
 19861 19960: gap of 100 bp  
 19961 20732: contig of 772 bp in length  
 20733 20832: gap of 100 bp  
 20833 21590: contig of 758 bp in length  
 21591 21690: gap of 100 bp  
 21691 22446: contig of 756 bp in length  
 22447 22546: gap of 100 bp  
 22547 23333: contig of 787 bp in length  
 23334 23433: gap of 100 bp  
 23434 24212: contig of 779 bp in length  
 24213 24312: gap of 100 bp  
 24313 25093: contig of 781 bp in length  
 25094 25193: gap of 100 bp  
 25194 25958: contig of 765 bp in length  
 25959 26058: gap of 100 bp  
 26059 26839: contig of 781 bp in length  
 26840 26939: gap of 100 bp  
 26940 27705: contig of 766 bp in length  
 27706 27805: gap of 100 bp  
 27806 28568: contig of 763 bp in length  
 28569 28668: gap of 100 bp  
 28669 29419: contig of 751 bp in length

29420 29519: gap of 100 bp  
 29520 30311: contig of 792 bp in length  
 30312 30411: gap of 100 bp  
 30412 31202: contig of 791 bp in length  
 31203 31302: gap of 100 bp  
 31303 32082: contig of 780 bp in length  
 32083 32182: gap of 100 bp  
 32183 32934: contig of 752 bp in length  
 32935 33034: gap of 100 bp  
 33035 33796: contig of 762 bp in length  
 33797 33896: gap of 100 bp  
 33897 34683: contig of 787 bp in length  
 34684 34783: gap of 100 bp  
 34784 35552: contig of 769 bp in length  
 35553 35652: gap of 100 bp  
 35653 36406: contig of 754 bp in length  
 36407 36506: gap of 100 bp  
 36507 37294: contig of 788 bp in length  
 37295 37394: gap of 100 bp  
 37395 38159: contig of 765 bp in length  
 38160 38259: gap of 100 bp  
 38260 39017: contig of 758 bp in length  
 39018 39117: gap of 100 bp  
 39118 39890: contig of 773 bp in length  
 39891 39990: gap of 100 bp  
 39991 40774: contig of 784 bp in length  
 40775 40874: gap of 100 bp  
 40875 41643: contig of 769 bp in length  
 41644 41743: gap of 100 bp  
 41744 42521: contig of 778 bp in length  
 42522 42621: gap of 100 bp  
 42622 43393: contig of 772 bp in length  
 43394 43493: gap of 100 bp  
 43494 44268: contig of 775 bp in length  
 44269 44368: gap of 100 bp  
 44369 45149: contig of 781 bp in length  
 45150 45249: gap of 100 bp  
 45250 46024: contig of 775 bp in length  
 46025 46124: gap of 100 bp  
 46125 46868: contig of 744 bp in length  
 46869 46968: gap of 100 bp  
 46969 47720: contig of 752 bp in length  
 47721 47820: gap of 100 bp  
 47821 48596: contig of 776 bp in length  
 48597 48696: gap of 100 bp  
 48697 49462: contig of 766 bp in length  
 49463 49562: gap of 100 bp  
 49563 50335: contig of 773 bp in length  
 50336 50435: gap of 100 bp  
 50436 51207: contig of 772 bp in length  
 51208 51307: gap of 100 bp  
 51308 52078: contig of 771 bp in length  
 52079 52178: gap of 100 bp  
 52179 52954: contig of 776 bp in length  
 52955 53054: gap of 100 bp  
 53055 53813: contig of 759 bp in length  
 53814 53913: gap of 100 bp  
 53914 54675: contig of 762 bp in length  
 54676 54775: gap of 100 bp  
 54776 55537: contig of 762 bp in length  
 55538 55637: gap of 100 bp  
 55638 56413: contig of 776 bp in length  
 56414 56513: gap of 100 bp  
 56514 57278: contig of 765 bp in length  
 57279 57378: gap of 100 bp  
 57379 58149: contig of 771 bp in length  
 58150 58249: gap of 100 bp  
 58250 59023: contig of 774 bp in length  
 59024 59123: gap of 100 bp  
 59124 59906: contig of 783 bp in length  
 59907 60006: gap of 100 bp  
 60007 60792: contig of 786 bp in length  
 60793 60892: gap of 100 bp



```

Db 26351 GATCAATAAAATCTGTAAATTTGAATGAGT 26381
|||||
RESULT 7
AK026889 2079 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ23236 fis, clone COL00725.
DEFINITION AK026889
ACCESSION AK026889
VERSION oligo capping; fis (full insert sequence).
KEYWORDS Homo sapiens cDNA to mRNA, clone_lib:COL clone:COL00725.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2079)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
DIRECT SUBMISSION
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leins.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source Location/Qualifiers
1. 2079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL00725"
/clone_lib="COL"
/tissue_type="colon"
/notes="cloning vector pME18SFL3"
304..681
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15584.1"
/db_xref="GI:10439855"
/translation="MDITKLIVQEKPKGSQRKRGHTASEDEQOWPEEKRLKEDI
LENEDEQNSPPKGRGRPPKPLGGTPEKPTMTKSKGSKKSGPPAPEEEEEER
QSGNTEQKSKQHRVSRRAQR"
BASE COUNT 759 a 354 c 339 g 627 t
ORIGIN

Query Match 8.4%; Score 441; DB 89; Length 2079;
Best Local Similarity 100.0%; Pred. No. 9.Be-227;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3689 gctcgaatggagaagcctagaggcaggaagaaaccccgctcacagacaggaggagaa 3748
|||||
Db 237 GTCCTGAATGGAGAGCCTAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 296
|||||
QY 3749 attagtgatgaatgactgaactaagtgtacagaaacacacacacacacacacacacacg 3808
|||||
Db 297 ATTAGTGATGGATGACTTGACTTAAGTTGTCAGGAACAGAAACCTTAAGGCAGTCAGCG 356
|||||
QY 3809 aagtcggaagagccatagcgttcagaatctgatgaacagcagtgccctgaggagaaa 3868
|||||
Db 357 AAGTCGGAAGAGGCCATACGCTTCAGATCTGATGAACAGCAGTCGCCTGAGGAGAA 416
|||||

```

```

QY 3869 gaggtcacaagaagatatattagaaaaatgaagatgaacagaatagtcgcccaaaaaaggg 3928
|||||
Db 417 GAGGCTCAAAAGAGATATATTAGAAAAATGAAGATGAACAGAATAGTCCGCCCAAAAAGGG 476
|||||
QY 3929 taaaagagccgacacacacacacacacacacacacacacacacacacacacacacacacac 3988
|||||
Db 477 TAAAAGAGGCCACACACACACACACACACACACACACACACACACACACACACACACACAT 536
|||||
QY 3989 gaaaacttctaaaaaaggaagcaaaaaaaatctggacctccagcaccacagagagagga 4048
|||||
Db 537 GAAAACCTCTAAAAAAGGAAGCAAAAAAAATCTGGACCTCCAGCACCAGAGAGGAGGA 596
|||||
QY 4049 agaagaagaagacacaaagtggaaatcacggaacagaaagtcacaaagcaaacacacagcagt 4108
|||||
Db 597 AGAAGAAGAAGACAAAGTGGAAATACGGAACAGAAAGTCCAAAAGCAACACAGCAGCGAGT 656
|||||
QY 4109 gtcaagagagacacacacacacacacacacacacacacacacacacacacacacacacac 4129
|||||
Db 657 GTCAAGGAGAGCACACAGCAGAG 677
|||||
RESULT 8
AC016449/c
LOCUS AC016449 72157 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016449
VERSION AC016449.2 GI:9104517
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 72157)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-45L14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72157)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6479175.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBK
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1189
Center clone name: 45_L14
-----
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for

```







[illegible]







||||| 3361 GAATGAAACTGTAAAGCAATGCAAGCAGAGCTCAATCAAGCTCTCTGGAAGATAA 3420  
QY 3577 agggagggcttgaagttctgaatgatcacagtgaaatgaagattacacaatctt 3636  
Db 3421 AGGGAGGCTTGATGCTTGAATGGATGACAGTGAATGAAGATTACAAATCTCT 3480  
QY 3637 cacctttgcggggaagaaagtgacaaagagagcactctgcttgaagtcgaat 3696  
Db 3481 CACCTTTGCGGGGAAAGGAGTGAAGAGAGAGAGGAGGAGGAGGAGGAGGAG 3540  
QY 3697 tggagagcctagaggcaggaagaaaccccgctcacagacagggaggagaaattaggta 3756  
Db 3541 TGAGAGGCTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3600  
QY 3757 tggatgactgactaagttggttacaggaacagaaacctaagagcagtcagcgaagtcga 3816  
Db 3601 TGGATGACTTGACTAAGTGGTACAGAGACAGAACTTAAGGCACTCAGCGAAGTCGGA 3660  
QY 3817 aaagagccatagcggcttcagaatctgtagacagcagtgccctgagggaagagggctca 3876  
Db 3661 AAAGAGGCTACCGCTTCAGAACTCTGATCAACAGCAGTGGCCTGAGGAGGAGGCTCA 3720  
QY 3877 aagaagatatattagaataatgaatgaacagaaatagtcgcgcaaaagggtaaaagag 3936  
Db 3721 AAGAAGATATATTAGAAATGAAGATGAACAGAAATAGTCCGCCAAAAGGGTAAAGAG 3780  
QY 3937 gccagaccacaaacctctgtgagggtacacaaagagagccaaacaaatgaacactt 3996  
Db 3781 GCGGACCACCAAACTCTGTGTGGAGGTACACCAAAAGAGAGGAGGAGGAGGAGG 3840  
QY 3997 ctaaaaaggaagcaaaaaactctggacctccagcaccagaggaggaggaggaagag 4056  
Db 3841 CTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3900  
QY 4057 aaagcaagtggaatacgggaacagagaagtcacaaagcaaaagcagccagtgtaagga 4116  
Db 3901 AAAGCAAAAGTGGAAATACGGAACAGAAAGTCCAAAAGCAAGCAGCAGGAGTCAAGGA 3960  
QY 4117 gacacagcagagagcagatctcctggaatctgtagtgaatgaatccacacagtcacac 4176  
Db 3961 GAGCAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4020  
QY 4177 cacagaaagcagaggaagaccatcaaaacgcacatccatccacacacacacacacac 4236  
Db 4021 CACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4080  
QY 4237 t 4237  
Db 4081 T 4081

RESULT 4  
LOCUS HSU050533 1852 bp mRNA PRI 27-NOV-1996  
DEFINITION Human BRCA2 region, mRNA sequence CG008.  
ACCESSION U050533  
VERSION U050533.1 GI:1531605  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1852)  
AUTHORS Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,  
Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,  
Farid,L., Frye,C., Hattier,T., Janek,T., Jiang,P., Kehrer,R.,  
Leibland,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y.,  
Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,  
Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,  
Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,  
Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,  
Tavtigian,S.V. and Simard,J.

TITLE Generation of an integrated transcription map of the BRCA2 region  
on chromosome 13q12-q13  
JOURNAL Genomics 36 (1), 86-99 (1996)  
MEDLINE 96411650  
REFERENCE 2 (bases 1 to 1852)  
AUTHORS Simard,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular  
Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,  
Quebec City, Quebec G1V 4G2, Canada  
FEATURES  
source  
1..1852  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13q12-q13"  
/note="CG008; DSEG numbers: D13S171 and D13S310"  
BASE COUNT 609 a 348 c 379 g 505 t 11 others  
ORIGIN

Query Match 21.4%; Score 1126; DB 97; Length 1852;  
Best Local Similarity 99.7%; Pred. NO. 0;  
Matches 1446; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1129 ttaaagtgaggtcacatgacccctgaggaagctattctctggtcaatgatcacttaatttgcgaag 1188  
Db 1 TTAAGCTGAGGTGACATGACCTTGAGGAGGCTATTAGACATGATGTTATTGTCGAATAG 60  
QY 1189 ttacagctctcaaaaggattctctctggtcaatgatcacttaatttgcgaag 1248  
Db 61 TTACAGCTCTAAAAAGGATATTCTTCTGTCATGATGATGTTTACTTAATTTTGTGAGAG 120  
QY 1249 agagaacattagacaacagatgagagtagcgaagagcgaagcagcagcagcagcagcagcagc 1308  
Db 121 AGAGAACATTAGACAACCGATGGAGAGTAGGCAAAAGCCATGATGGGACTTGGCCAAA 180  
QY 1309 ttataagaataatgctttcacagtcagcagcagcagcagcagcagcagcagcagcagcagcagc 1368  
Db 181 TTTATAGAATAATGCTTTTACAGTCAGCAGCTGGAAAGATGCTGCACAAACAGATAGCAT 240  
QY 1369 ggaacaaagacaaattgctacatatattatacaaaatagttatgatgactgactctg 1428  
Db 241 GGATCAAGACAAATTTGCTACATATATATATCAAAATAGTATTGATGATCGACTACTTG 300  
QY 1429 ttgaacggtcttctgctcaatcagctgctcctcacaatttagaaactcagacagcagatga 1488  
Db 301 TTGAACGGATCTTGTCTCAATACATGCTTCTCACAATTTAGAAACTACAGACGGATGA 360  
QY 1489 aatgcttattacttctgctgacacactggatttaaatgctgtgaagcatttgaatgaag 1548  
Db 361 AATGCTTATATTACTTGTATGTCACACTGGATTTAAATGCTGTGAAGCATTTGAATGAA 420  
QY 1549 tgtgaaatgtcaaaatctgctccgacacacaaagatttgcctgactgactgaagc 1608  
Db 421 TGTGGAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTGACTGATTAAAG 480  
QY 1609 aacccaaacagatgacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1668  
Db 481 AACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGTGTTATTACAAGAA 540  
QY 1669 attactctgactctggttaaggctcaggatttcattgaagaataatcacacaggtgttagaag 1728  
Db 541 ATTACTGATCTCTGTTAAAGGCTCAGGATTTTCATGAAGAAATTCACACAGGCTGTGAAG 600  
QY 1729 atgataagaaataaagaacagattagaagtagctttagtccaacactgctcctgaagc 1788  
Db 601 ATGATGAGAAATTAAGAAAGCAGTGTAGAGTACTTGTAGTCAACATGCTCTCTGCAAGC 660  
QY 1789 aggctgaaggttctgctgctgaaataactaaagaagttgggcaaccccaacacagcctcaca 1848  
Db 661 AGGCTGAAGGTTGTGCTGCTGAATTAACTAAGAAGTTGGGCAACCCCAACACACGCTTACA 720



QY 1357 aacagatagcatgatacaagacaaattgctacatatattatcaaaaatagattgatg 1416  
DB 1201 AACAGATAGCATGATCAAGACAAATGCTACATATATATCAAAATAGTATTGATG 1260  
QY 1417 atcgactactgttgaaagcgtatcttctcaatacatagttcccaaatttagaacta 1476  
DB 1261 ATCGACTACTGTGTGAACGGATCTTTGCTCAATACATGTTCCCTCACAATTTAGAAACTA 1320  
QY 1477 cagaacggatgaaatgattattacttattgtagccacacgtgattttaaagtctgtgaaag 1536  
DB 1321 CAGAACGGATGAAATGCTTATATTTACTTGTATGCGCACACTGGATTTAAATGCTGTGAAG 1380  
QY 1537 cattgaatgaaatggaatgtcaaaatctgctccacacatcaagtaagagatttgcctg 1596  
DB 1381 CATTCATGAAATGTGGAATGTCAAAATCTGCTCCGACATCAAGTAAGATTTGCTTG 1440  
QY 1597 acttgattaaagcaaccacaaacagatgccagtgccaagggccatatattcaaaagtgtg 1656  
DB 1441 ACTTGATTAAAGCAACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGG 1500  
QY 1657 ttattacaagaaattaccctgatactctgtaaggctcaggatttcaatgaagaaattcacac 1716  
DB 1501 TTATTTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTTCATGAAGAAATTCACAC 1560  
QY 1717 aggtgttagaagatgagagaaataagaagcaggttagaagtagtctgttagtccaacat 1776  
DB 1561 AGGTGTTAGAAGATGATGAGAAATAGAAGACAGTTAGAAAGTACTTGTAGTCCCAACAT 1620  
QY 1777 gctcctgaagcagcgtgaaggtgtgtgctgtaaatcaactaagaagttgggcaacccca 1836  
DB 1621 GCTCCTCAAGCAGCGTGAAGTTGTGTGCGTGAATTAACATAAGAACTGGGGCAACCCCA 1680  
QY 1837 aacagctacaatcctttcctggaatgatcaagtttctcttgagagagatacactg 1896  
DB 1681 AACAGCCTACAAATPCTTCTCTGGAATGATCAAGTTTCTCTTGAGAGGATAGCACCTG 1740  
QY 1897 tgcacatagatcacgaatctatcagtcctcttattaacaagtgaaacaaatcaatagatg 1956  
DB 1741 TGCACATAGATACCGATCTATCACTGCTCTTATTAACAAGTGAACAAATCAATAGATG 1800  
QY 1957 gaacagcagatgataagatgaggtgttccaaactgatcaagccatcagagcaggtcttg 2016  
DB 1801 GAACAGCAGATGATGAAGATGAGGCTGTTCAACTGATCAAGCCATCAGAGCAGGTCTTG 1860  
QY 2017 aactgcttaaggctactctcattacacatcccatctcatttctcattctctgtaaacatttg 2076  
DB 1861 AACTGCTTAAGGTACTCTCATTTACATCCCATCTCATTTCAATTTCTGCTGAACATTTG 1920  
QY 2077 aatcattactggctgtgctgaatgtagatgaaagtagcagaagctgcactacaac 2136  
DB 1921 AATCATTTACTGGCTTGCTGAAATGGATGATGAAAGTAGCAAGAGCTGCATACAAA 1980  
QY 2137 ttttcaaaaacacaggaagcaaaattgaagaggattttccacacatcagatcagccttgc 2196  
DB 1981 TTTTCAAAAACACAGGAAGCAAAATTCAGAGGATTTTCCACACATCAGATCAGCCTTGC 2040  
QY 2197 ttcctgttttacatcaacaaatctaaaaaagagcccccccgctcaagccaaatattgcattc 2256  
DB 2041 TTCTGTGTTTACATCACAAATCTAAAAAAGGACCCCGCCGCTCAAGCCAAAATATGCCATTC 2100  
QY 2257 attgatccatgcgatatcttctagtaaaagaccaggtttgcacagatatattgagcctc 2316  
DB 2101 ATTGTATCCATGCGCATATTTTCTAGTAAAGAGACCCCGAGTTTGCACAGATATTTGAGCCTC 2160  
QY 2317 tgcataagagcctagatcccaagcaacctgggaactctcaataacacattggttactattg 2376  
DB 2161 TGCATAAGAGCCTAGATCCCAAGCAACCTGGAAACATCTCATAAACACCATTTGTTACTATTG 2220  
QY 2377 gtcattatgctcctctgcaactgatacaatttgcctcttgaaatctcttggaagtagcta 2436  
DB 2221 GTCATATTTGCTCTCTGACCTGATCAATTTGTGTCTCTTTGAAATCTTTTGTGTAGCTA 2280

QY 2437 ctttcatttgaaagatcttctcattgatcggtccagggaagaaagacaactaaac 2496  
DB 2281 CTTTCATTGTTGAAGATCTTCTCATGATCGCTTCCAGGGAAAAAGACAACATAAC 2340  
QY 2497 ttggggttccagatgaagaagtatctcctgagacaaatgggtcaaaatctcaggtctattaaaa 2556  
DB 2341 TTTGGGTTCCAGATGAAGAAGTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAAA 2400  
QY 2557 tgatggttccatggctacttgggaatgaaaaataatcacagtaaatcacaggaacttctaacct 2616  
DB 2401 TGATGGTTCGATGCTACTTTGGAATGAAAAATATACAGTAAATCAGGAACCTTCTACT 2460  
QY 2617 taagattgcttaacaacaatattgcatagtgagacttgacagacagaggggaaaaatta 2676  
DB 2461 TAAGATTGCTTAACAACAATATTGCTAGTATGATGGAGACTTGCAGAGACAGGGGAAAAATTA 2520  
QY 2677 gtaaacagatatagtcacgtctgagacttgcgtggagtgctattgtgaagctggcac 2736  
DB 2521 GTAAACCAGATATGTCACGCTGAGACTTGTCTGGGAGTGTCTATTGTGAAGCTGGCAC 2580  
QY 2737 aagaacccctgttaaccatgaaataatcacattagaaacaatatcaagctatgtgcatagcta 2796  
DB 2581 AAGAACCCTGTTACCATGAAATCATCATTTAGAACAAATATCAGCTATGTGCAATTAGCTA 2640  
QY 2797 tcaacgatgaatgctatacaagtaagacaagtgtttgccagaaacttccacaagccttt 2856  
DB 2641 TCAACGATGATGCTATCAAGTAAGACAAGTGTGTTGCCAGAACTTCACAAAGCCCTTT 2700  
QY 2857 ccggtttacggcttccacttgagtataggcaatctctgcccctttgtgcaaaaagactcg 2916  
DB 2701 CCCGTTTACGGCTCCACTTTCAGTATATGGAATCTGTGCCCTTTGTGCCAAAAGATCCTG 2760  
QY 2917 taaaggagaagagctcatcgttagggcaatgtttgggtgaaaaataataaatgaaggcggg 2976  
DB 2761 TAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTGGTGAAAAATATAAATGTAGAGCGGG 2820  
QY 2977 agtatctgaagcagcagcagcagctgttagtgaataattattctctctctaccagagtag 3036  
DB 2821 AGTATCTGAAGCAGCATGCGCTGTAGTGAATAATTTATTGCTCTCTCTACCCAGATG 2880  
QY 3037 ttgttccatatacaattcaccttttggcactgaccagagattatgtcaaaagtcaggata 3096  
DB 2881 TTGTTCCATATACAAATTCACCTTTTGGCACATGACCCAGATTTATGTCAAAGTACAGGATA 2940  
QY 3097 ttgaacaacttaagatgttaagaatgtcttggtttcttggaataattatagtagcta 3156  
DB 2941 TTGAACAACCTTAAAGATGTTAAAGAAATGCTTTGGTTGTTGTTCTGGAAATATTAATGGCTA 3000  
QY 3157 aaaaatgaaaaataacagtcacgcttttatcagaagaatggtagaaaaattataaacaacaa 3216  
DB 3001 AAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAACAACAA 3060  
QY 3217 aagatgccccaggaccagatgatgcaaaaaatgaatgaaaaactgtacactgtgtgtgatg 3276  
DB 3061 AAGATGCCCAAGGACCAGATGATGCAAAATGAATGAAAACTGTACACTGTGTGTGATG 3120  
QY 3277 ttggcatgaatatcatcatgctcaaaagagtaactacacagtttggaaatctcctcaaaagacc 3336  
DB 3121 TTGCGATGAATATCATCATGTCAAAAGAGTACTACATACAGTTTGGAAATCTCCTTAAAGACC 3180  
QY 3337 cggctactaccagctcgtttcttctcactcaacctgacaagaatttcagtaacaccccaaaatt 3396  
DB 3181 CGGTACTACCAAGCTCGTTTCTTCACTCAACCTGACAGAATTTTCAGTAACACCAAAATTT 3240  
QY 3397 atctgctccttgaaatgaaatcatttttctcactcctggaaaaacccctaaaaacacaaatgttc 3456  
DB 3241 ATCTGCTCTCTGAAATGAATGAAATCATTTTCACTCTCTGAAAAACCTTAAAAACAACCAATGTT 3300  
QY 3457 tagagctgtttaacaagccacttcatcagcaggaagcaatctcagacccaataatcatcac 3516  
DB 3301 TAGAGCTGTTTAAACAAGCCACTTTTCATCAGCAGGCAAGCAATCTCAGACCAAAATCATCAC 3360  
QY 3517 gaaatgaaaactgtaagcaatgcaagcagcagctcaaatcccaagctcctcctctggaagaaataa 3576



[illegible]



```
Db 2282 AAAAGGACCCCGCTCAAGCCAAATATGCATTCATTTGTATCCATCGCATATTTTCTA 2341
Qy 2281 gtaagagaccaggtttgcacagattttgagctctgcataaagagcctagatccaagca 2340
Db 2342 GTAAGGACCCAGTGTGCACAGATATTGAGCCCTCGCATAGAGGCTAGATCCAGCA 2401
Qy 2341 acctggaacatctcataacaccattggttactattgctcatattgctctccttgcacctg 2400
Db 2402 ACCTGGAAACATCTCATACACCATTTGTTACTATTGGTCATATTGCTCTCTGCACCTG 2461
Qy 2401 atcaattgtctcctctggaaactcttgggttagctacttccattgtgaaagattcttca 2460
Db 2462 ATCAATTTGCTGCTCTTTGAAATCTTTGGTAGCTACTTTTCATTTGTAAGATCTTCTCA 2521
Qy 2461 tgaatgatcggcttcagggaaagaaagcaactaaactttgggttccagatgaagaatat 2520
Db 2522 TGAATGATCGGCTTCCAGGGAAAAGAACACTAACTTTTGGGTTCCAGATGAAGAATAT 2581
Qy 2521 ctctcgagacaattggtcaaaattcaggctatttaaaatgtagtctcgatggctacttggaa 2580
Db 2582 CTCCTGAGACAATGGTCAAAATTCAGGCTATTAAATGATGGTTTCGATGGCTACTTGGAA 2641
Qy 2581 tgaataataatcacagtaaatcaggaactctaccttaagattgctaaacaataattgc 2640
Db 2642 TGAATAATAATCACAGTAAATCAGGAACCTTCTACCTTTAAGATTGCTAAACAATAATTGC 2701
Qy 2641 atagtgtgagacttgacagaaacaggggaaaaattagtaaacagatatgtcacgcttga 2700
Db 2702 ATAGTGATGGAGACTTGACAGAAACAGGGGAAAATTTAGTAAACAGATGATGTCAGCTGA 2761
Qy 2701 gacttgcctggtgagtgctattgtaagctggcacagaacacctgtttacattgaatacca 2760
Db 2762 GACTTGTCTGGGAGTCTATTGTGAAGCTGGCACAGAACCCTGTTACCATGAATCA 2821
Qy 2761 tcacattagaacaatacagctatgtgtaattagctatcaacgatagaatgtctatcaagtaa 2820
Db 2822 TCACATTAGAACAATATCAGCTATGTGCTATGTAGCTATCAACGATGAATGCTATCAAGTAA 2881
Qy 2821 gacaagtgttgcagagaacttcacaaagcctttcccggtttacggtcttccacttgagt 2880
Db 2882 GACAAGTGTGTCAGAGAACTTCACAAAGGCCCTTCCCGTTTCCGCTTCCACTTGTAGT 2941
Qy 2881 atatggcaactgtgccctttgtgcaaaagatcctgttaaaagagagagagctcatgcta 2940
Db 2942 ATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGAGAGAGAGCTCATGCTA 3001
Qy 2941 ggcattgttggtaaaaaataaataatgaagcgggagatctcgaagcagcatgcagctg 3000
Db 3002 GGCAATGTTTGGTGAAAATATAATGTAAAGCGGGAGTATCTGAAGCAGCATGCAGCTG 3061
Qy 3001 ttatgaaaaattattgtctcttaccagagtagtattgtttccatatatacaattcaccttt 3060
Db 3062 TTATGTAAAAATATTGTCTCTTCTACAGAGTATGTTGTTCCATATACAAATTCCCTTT 3121
Qy 3061 tggcacatgacccagattattgcaagtcacaggatattgaaacaacttaaaagtgttaag 3120
Db 3122 TGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACCTTAAAGATTTAAAG 3181
Qy 3121 aatgtcttggttgttctcgaaattattaatggttaaaaaatgaaataaataacagtcacgtt 3180
Db 3182 AATGCTTTGTTGTTTCTGTAATATTAAATGCGTAAAAATGAAAATAACACAGTCACGCTT 3241
Qy 3181 ttatcagaagaatgtagaataatttaacaaacaaagatgccaaagaccagcatgatg 3240
Db 3242 TTATCAGAAAGATGTAGAAAATATTAAACAAACAAAGATGCCAAGGACGATGATG 3301
Qy 3241 caaaaaatgaataaaaaactgtacactgtgtgtgattggtgcatgaataatcatcatgtcaa 3300
Db 3302 CAAAAATCAATGAATAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAA 3361
Qy 3301 agagtactacatacagatttgaatctctcaaaagaccgggtactaccagctcgtttcttca 3360
Db 3362 AGAGTACTACATACAGTTTGGAAATCTCTTAAAGACCCGGTACTACCAGCTCGTTTCTTCA 3421
```

## RESULT 3

```
AB023196
LOCUS AB023196 5110 bp mRNA PRI 16-JUN-1999
DEFINITION Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION AB023196
VERSION AB023196.1 GI:4589601
KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBBluescriptII
SOURCE SK plus clone:hj07056.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



QY 61 ctgtcatggtctatccaagactagagccaatgatggaaaaattacatatcccgctgggg 120  
DB 122 CTGTCTATGGCTCAATTCAAGACTAGGACCAATGATGGAATAATTACATATCCGCTGGG 181  
QY 121 tcaagaaatcatcagataaataatcctaaagagagatggtgagacattaaagatggtg 180  
DB 182 TCAAGGAATAATCAGATAAATAATCTTAAAGAGGAGATGGTGACAGATTAAGATGGTG 241  
QY 181 tgaactttttatggatggaccagactctgaagaagaaaaaggagctttatttaaacc 240  
DB 242 TGAATACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAAGGAGCTTTATTAAACC 301  
QY 241 tagctttacatctgtctcagatattttttctcaagatccctgtaagatggttcgcttac 300  
DB 302 TAGCTTTACATCTGTCTCAGATTTTCTCAAGCATCTGATAAAGATGTTTCGCTTAC 361  
QY 301 tggtagctgctgctctgctgatatttcaaggatttatgctcctgaagctccctacacat 360  
DB 362 TGGTAGCCTGCTGCTGTGATATTTTCAGGATTTATGCTCTGGAAGCTCTTACACAT 421  
QY 361 ccctgataaaactaaaggatatatttattgtttatacaagacaggttgaagggtctaggg 420  
DB 422 CCCTGATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGCTAGAGG 481  
QY 421 atacaagagccacaaattcaataggtattttttatttacttgagaacattgctgggtca 480  
DB 482 ATACAAGAGCCACAAATTCATAGGTATTTTATTACTTGAAGAATTTGCTTGGGTCA 541  
QY 481 agtcataacatattgcttgagttagaagatagcaatgaattttccaccagctataca 540  
DB 542 AGTCATATACATATGCTTTGAGTTTGAAGATAGCAATGAAATTTTCAACCACCTATACA 601  
QY 541 gaaccttatttcagttataaaacattggccacaaatcagaagaatcccatatgacatggtag 600  
DB 602 GAACCTTATTTTCAGTTATAAACAATGGCCACAATCAGAAGTCCATATGCACATGGTAG 661  
QY 601 acccttagctctattatttgaagggtgatacagtgctcagggagcttttgatacgg 660  
DB 662 ACCTTATGAGCTCTATTTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGG 721  
QY 661 ttttagtaaatcgtgctacgtcctataagaatttaacaagcaagcatatgatttggcaa 720  
DB 722 TTTTGTAAATCTGGTACCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTTGGCAA 781  
QY 721 aggtcttactgaagagagacgtccaagctattgagccatatattaccacttttttaatc 780  
DB 782 AGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTAAATC 841  
QY 781 aggtctctgattggtgaaacacatcatcagcatttgcagagcatgctttgacttaa 840  
DB 842 AGGTTCTGATGCTTGGGAACAATCTATCAGCGATTGTGTCAGAGCATGCTTTTGACTTAA 901  
QY 841 ttttgagctctacaatttgaagtcattgctgctctctgcttttaccaccagcttgaat 900  
DB 902 TTTTGGAGCTCTACATATTGATAGTCATTTGCTGCTCTGTTTACCCAGCTTGAAT 961  
QY 901 ttaaaataaagagcaatgataa tgaaggagcgctacaagttgttaaactactgccaataa 960  
DB 962 TTAATAATAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTAAACTACTGGCAAAA 1021  
QY 961 tgtttgggcaagaggttcagaattggcttctcaaaacagccactttggcagtgctact 1020  
DB 1022 TGTTTGGGCAAGGATTCAGAAATTTGCTTCTCAAAACAGGCCACTTTGGCAGTGTACT 1081  
QY 1021 tggcgagggtttaaataatccatgtaccatccgcttgaattgtgaaatttgcagcc 1080  
DB 1082 TGGCAGGTTTANTGATATCCATGTACCAATCCGCTGGAATGTGCAATTTTGTAGCC 1141  
QY 1081 attgtctcatgaaccatctcgtatttagcaaaagacttaacagagtatcttaaaagtgggt 1140  
DB 1142 ATTGTCTCATGAACCATCTCTGATTTAGCAAAAAGACTTAACAGAGTATCTTAAAGTAGGT 1201  
QY 1141 cacatgacctgaggagactatttagacatgatgttatttggtaagttacagctgcta 1200

DB 1202 CACATGACCCCTGAGGAGCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGCTA 1261  
QY 1201 aaaagatatcttcttggtcaatgatcactactacttaatttctgagagagagaacattag 1260  
DB 1262 AAAAGGATATCTTCTGGTCAATGATCCTTACTTAATTTTGTGAGAGAGAAACATTAG 1321  
QY 1261 acaaacgatggagagtagtcgcaaaagaagccatgatgggacttgcaccaattataagaat 1320  
DB 1322 ACAACGATGAGAGTAGCGCAAGAGCCATGATGGGACTTGCCCAATTTATAAGAAAT 1381  
QY 1321 atgctttcaagtcagcagctggaaaaagatgctgcaaaacagatagcatgatacaaaaca 1380  
DB 1382 ATGCTTTACAGTCAGCAGCTGSAAGAGATGCTGCAAAACAGATAGCATGGATCAAGACA 1441  
QY 1381 aattgctacatatattatcaaaatagttatgatcagctacttctgtgaacggatct 1440  
DB 1442 AATTGCTACATATATATTATCAAAATAGTATTGATGATCAGCTACTTGTGTAACGGATCT 1501  
QY 1441 ttgctcaatcacatgggttctctccacaatttagaaactacagacacggatgaaatgcttatct 1500  
DB 1502 TTGCTCAATACATGGTTCTCTACAATTTAGAAACTACAGAACGGATGAAATGCTTATATT 1561  
QY 1501 acttgatgccacacitggattttaaagtctgaaagcatlgaatgaaatgtgaaatgctc 1560  
DB 1562 ACTGTATGCCACACTGGATTTAAATGCTGTGAAGCATTTGAATGAAATGTGAAATGTC 1621  
QY 1561 aaaaactgctccgacatcaagtaaaagattgcttgaacttgattaaagcaaccccaaacag 1620  
DB 1622 AANAATCTGCTCCACATCAAGTAAAGGATTTGCTTGACTTGTATTAGCAACCCCAACAG 1681  
QY 1621 atgccagtgtaaggccatattttcaaaagtgatgggttattcaagaataattacctgctc 1680  
DB 1682 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATC 1741  
QY 1681 ctggtaaggctcaggatttcaatgaagaatttcacacaggtgttagaaga tgaagagaaaa 1740  
DB 1742 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTTACACAGGTGTTAGAAAGATGATGAGAAA 1801  
QY 1741 taaagaagcagttlagaagcttctgttagtccaaacatgctcctcgaagcaggtgaaagtt 1800  
DB 1802 TAGAANAAGCTTTAGAAAGTACTTGTTAGTCCCAATGCTCTGCAAGCAGCTGAAAGTT 1861  
QY 1801 gtgtcggtgaaa taaactaagaagttggggcaaccccccaacagcctcaaaa tcccttccctg 1860  
DB 1862 GTGTGGCTGAAATAAATTAAGAAGTTGGGCAACCCCAACAGCCTTACAAATCTTTCCTGG 1921  
QY 1861 aagatgataagtttctcttgagagatagcacctgtgcaacatagataccgaatctatca 1920  
DB 1922 AAATGATCAAGTTTCTCTTGGGAGGATAGCACCCTGTGACATAGATACCGAATCTATCA 1981  
QY 1921 gtgctcttataaacaagtgaacaaatcaatagatggaacagcagatgataagagatgagg 1980  
DB 1982 GTGCTCTTATTAAACAAGTGAACAATCAATAGATGAACAGCAGATGATGAAGATGAGG 2041  
QY 1981 gtgtccaaactgataagccatcagagcaggtcttgaactgcttaaggtactctcatcta 2040  
DB 2042 GTGTTTCAACTGTATCAAGCCATCAGAGCAGGTCTTGAACCTGCTTAAAGGTACTCTCATTTA 2101  
QY 2041 cacatccatctcatttctctgctgaacatttgaatcattactggtctgctgaaaa 2100  
DB 2102 CACATCCCATCTCTATTTCTGCTGAAACATTTGAATCATTTACTTGGCTTGTCTGAAA 2161  
QY 2101 tggatgatgaaaaagtagcagaagctgactacataaaattttcaaaaaacacaggaagcaaaa 2160  
DB 2162 TGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAA 2221  
QY 2161 tgaagagagattttccacacatcagatcagccttgcctctgttttaca tcaacaatctta 2220  
DB 2222 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGGCTTCTGTTTACATCACAATACTA 2281  
QY 2221 aaaaagacccccctcaagccaaatagcatttctgataccatgcatatttttcta 2280







```
QY 4081 agaagtcacaaacacacagaccgagtgtaagagagacacagcagagagacgaatctc 4140
DB 4081 AGAAGTCCAAAACAAACACGACCGAGTGTCAAGAGAGACACAGACGAGAGAAATCTT 4140
QY 4141 ctgaatctagtgcaatgaatcacacagctccacacacacagaaagacagagagacac 4200
DB 4141 CTGAATCTAGTGAATGAATCCACACAGTCCACACACACAGAAAGACGAGGAGACCAT 4200
QY 4201 caaaacgcacatcacacacacacacacacacacacacacacacacacacacacacac 4260
DB 4201 CAAAACGCCATCACCATCACACACACACACACACACACACACACACACACACACATTC 4260
QY 4261 aaaccaattcaaatattttgcaaaagtcttaaatattgtaaacacacacacacacacac 4320
DB 4261 AAACCAATTCAAATATTTTGCAAAAGTTCCTAAATTTGTAAACATACATATTGCTGA 4320
QY 4321 tttaaatccatattttgacccattacactaggtacgagcgcgagtgctaaagagga 4380
DB 4321 TTTAAATTCATATATTTAGCCCATTTACACTAGGTACGCGCGCGAGGTGCTAAAGGGA 4380
QY 4381 agggcgatgaacaaatgaattaactttctctgtgaaagctttggaaaaatctttt 4440
DB 4381 ACGGCATGAACAAATGTAATAAATTAACCTCTCTGTAAGACTTTGGAAAAATCTTTT 4440
QY 4441 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 4500
DB 4441 TTTTATTTTGTGTCAGCTTGAGGCTTGAGGCTGAATAAGGCTTTGATGCACAAAATGG 4500
QY 4501 gactgctgaagtgacagtgacacttacttctgtgacccacacacacacacacacacacac 4560
DB 4501 GACTGCTGAAGTGGACAGTGGACCTTACTTTTGGTGACCCCATACATTTGGTGACACA 4560
QY 4561 tgccttagccatacacatggaacattgactatggagcttctgtgaaagtgaatgtgcga 4620
DB 4561 TGCTTAGCCATACACATGCTAATGCTGCTGTAAGTCTTGTAAGTCTAATGTGCGA 4620
QY 4621 tggctatgacacataaagaagaacttgtaaatctcttttttttttttttttttttttt 4680
DB 4621 TGCTATGTAGACATAAAGAAGAACTTGAATAATCTTTTCTTTTCTTTTAAATGCTTC 4680
QY 4681 tgattctgaagtctgtatagcttttctctgagctttaaactgacagtagccagctg 4740
DB 4681 TGATTTCTGAAGTCTGTATAGCTTTTATCTGCGGCTTTAACTGACAGTACCCGACTG 4740
QY 4741 tttatggatctattgatttgaagaagtgttaggatagatctttaaagcagtaactctgt 4800
DB 4741 TTTATTTGGATCTATTGATTGAAAAGAAATTTGTTAGGATAGATCTTTAAGCAGTAATCTGT 4800
QY 4801 cagtggtttattgtatttctgcaatttactgtgaaacacacacacacacacacacacacac 4860
DB 4801 CAGTGTGTTGATTGTTTCTGCAATTTTACTGTGAAAAAAATTTGTTTCAACAAT 4860
QY 4861 tgggtcattttcttgatgctactattgttgagaggttaaatggctctctcccttctg 4920
DB 4861 TGGTGTCAATTTCTGATGCTACTATTGTTGGAGAGTTAAATGGTCTCTTCCCTTTGTC 4920
QY 4921 tatcttacctagtggttactctctggcacccttaattcttcagaggtgctaaattgtctgc 4980
DB 4921 TATCTTACCTAGTGTGTTTACTCTGGCACCCCTTAATCTTTCAGAGGTGCTAAATTTGCTGC 4980
QY 4981 cattacacagaagagtgctctgtagagagacacacacacacacacacacacacacacacac 5040
DB 4981 CATTAACAGAGAGTGCCTCTGATAGAGGACACACACACACACACACACACACACATCTCT 5040
QY 5041 gaagtcttggattactttacacactcagttattgttgcacagaatttcttggccttcc 5100
DB 5041 GAAGTCTTTGGATTACTTTACACCTCAGTATGATTGTTGTTCCCGAGAAATTTCTGCGCCTTTC 5100
QY 5101 atggcaatgaaattttaagaagaagatttaaatattttaaagagtggtgta 5160
DB 5101 ATGGCAATGAAAAATTTTAAGAAGAAAGATTTAAAGTATTTTAAATTTTAAAGAGTGTGTTA 5160
QY 5161 taaaaataatgtactgaattctttatcccatctttatcatcacttcttcagtttatttaata 5220
```

```
DB 5161 TAAAAATAGTACTGAATCTTTTATCCCATTTTATCATCCTTTTATCAGTTTTTATTATCTA 5220
QY 5221 ctgtatcaataaaattctgttaatttgaatgagt 5253
DB 5221 CTGTATCAATAAAATTTCTGTAAATTTGAATGAGT 5253

RESULT 2
HS26H231 7444 bp mRNA PRI 25-JAN-2000
LOCUS Novel human gene mapping to chromosome 13.
DEFINITION AL137201
ACCESSION AL137201.1 GI:6759511
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Rhodes,S. and Huckle,E.
Direct Submission
Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk
This cDNA sequence was assembled from public domain ESTs and single
pass sequencing reads from expressed DNA templates, aligned to the
genomic DNA sequence from the bacterial clones 26H23 (284467),
267P19 (275889) and 49J10 (284572).
The EST sequences listed match this sequence with an identity of at
least 95% between the coordinates shown.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene
Sanger Centre name : 26H23.Cl3.1.

FEATURES
Location/Qualifiers
1..7444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12.3"
1..107
/number=1
108..234
/number=2
127..4470
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB69911.1"
/db_xref="GI:6759512"
/translation="MAHSKTFTNDGKITYPGCVKEISDKISKKEEMVRLKLVKVFMD
MQDSEEEKELYLNALHLASDFFLKHPDKDRLVLVACCLADIFRIYAEAPYTSDDK
LKDFMFITRQUGLEDTKSPQFNRYFYLLLENIAWVKSYNICFELEDSNEIFQIYRT
LFSVINGHNRQVHMVMDLMSIIICEGTVSOELDTVLNLVPAHNLNKAQYDLA
KALLKRTAQAIIEPYITNFFNOVLMLGKTSISDLSEHFVLDLILELYNIDSHLLSLVPO
LEFLKLSNDNEERLQVLLAKMFGAKDSLASQNSHKPLQCVLYGRFNDIHPVIRLECY
KFASCHLMNHPDLAKDLEYLKRSHDPEAIRHDVIVSVITAAKDKILLVNDHLLNF
VREPLDKRVRVRKEAMGLAQIYKYLQSAQDAKAAQIAIMDKDKLLHYIQNSID
DRLLVERIAQYVMPHNLETTERMKLYLYLATLDLNAVKNMKNKONLRLHQVD
LDDLKPKTDASVKAIFSKVNVITNPLPGKAQDMFKFTQVLEDDDKIRKQLEVL
VSPCTSCQAEGCVREITKLGKNPQPTNPFLMIKFLERLAPHDIFESALIKQ
VNSIDGTADEDEGVPTDQIARAGLELLKVLSTPHISFSAETFESLACLKMDDE
KVAEALQIIFKNGTSKIBEDPHIRSAIIPVLHHSKKGPPQAQYAIHICAIIFSSK
ETQPAQIFEPHLHKSLOPSNLEHLITPLVTIGHIALLPDQFAAPLKSUVATFVLDLL
MNDRLPGKKTTLKLPVDEEVSPEPTVKVQIAIKMVRWLGLGMKNHKSSTSTRLLT
ILHSDGLDTEQKISKPDMSRLRGAISAIVKLAQPCVYHEITILEQYQCALANDE
CYQVQVPAQLKHLGSLRLPLEYMAICACAKDPVKERRAHQVCKLWFLVETILMA
LKQHAASVEKLLSLPEYVVPYTHLLAHPDYVKVQDIEQLKDYKQCLWFLVETILMA
KNENSHAFIRKWNENIKQTKDAQDPDAKMKELTYTVCDAVMNIIIMSKSTYSLESP
KDPVLPAFPFTQPKDNSTNKNLPPPEMKSFPTPKPTTNVIGAVNRPKLSLSSAGKQSQ
TKSSMETVSNASSSSNPSPGRIRKGLSDSMHSENEDEYTMSSPLPKGKSKDRDSD
DLVRSELEKPRGRKKTPTVEQEKLGMDLTLKVOEQPKGSGRKRKRGHTASEDEQ
OWPEERLKDILENEDQNSPPKGRKRPKPLGGGTGPKPEPTMKTSGKSKKSG
PPAPEEEEEERQSGNTQKSKSKOHRVSRRAQRAESPESAIESTOSTPOKGRGP
SKTSPSPQPKNVRVGRSKQATKENDSSEEVDFQSGSPVDDIIPQETEEVEESTVN
VRRSAKRERR"
```



1861 AAATGATCAAGTTTCTTGGAGAGGATAGCACCTGTCACATAGATACCGAATCTATCA 1920  
1921 gtgccttattaaacaagtaacaaaatacaatagatgaacagagatgatgaacatcagg 1980  
1921 GTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGACAGATGATGAAGATGAGG 1980  
1981 gtgttccaactgatcaagccatcagacagagctcttgtaactgcttaagtgactctcattta 2040  
1981 GTGTTCCAATGATCAAGCCATCAGACAGAGCTCTTGAACCTTGAAGGTACTCTCAATTA 2040  
2041 cacatcccatctcatttcattctgctgaaacatttgaaatcattactgctgctgctgaa 2100  
2041 CACATCCCATCTCATTTCTGCTGAAACATTTGAATCATTTACTGGCCCTGCTGAAAA 2100  
2101 tggatgatgaaaaagtagcagaagctcactacaaattttcaaaaacacaggaagcaaaa 2160  
2101 TGGATGATGAAAAAGTAGCAGAAGCTGCACCTACAAAATTTTCAAAAACACAGGAAGCAAAA 2160  
2161 ttgaaggagattttccacacatcagatcagccttgcttgcctgcttttaccatcacaaatcta 2220  
2161 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCGTGTTTACATCACAAATCTA 2220  
2221 aaaaagaccccccgctgaagccaaaatgcccatttcatttgcattatccatgcgatatcttcta 2280  
2221 AAAAAGGACCCCCCGTCAAGCCAAATATGCCATTTCAATTTGATCCATCGCATATTTTCTA 2280  
2281 gtaagagacccagtttgcacagatatttgacctctgcataagagcctagatcccaagca 2340  
2281 GTAAGAGGCCAGTTTGCACAGATATTGAGCCTCTGCATAAAGAGCCTAGATCCAAAGCA 2340  
2341 acctggaacatctcataaacaccattggttactattggtcattatgctctcctctgcacctg 2400  
2341 ACCTTGGAAACATCTCATTAACACCATTTGTTACTATTGTTGTCATATTGCTCTCCCTGCACCTG 2400  
2401 atcaatttgcctccttggaatacttgggtagtaacttcaatttgcataagagatcttctca 2460  
2401 ATCAATTTGCTGCTTCTTGAATCTTTTGGTAGCTACTTTTCATTTGTAAGATCTTCTCA 2460  
2461 tgaatgatcgcttccagggaagaagcaactaaactttgggttccagatgaagaagtat 2520  
2461 TGAATGATCGGCTTCCAGGGAAGAAGCACTAAACTTTTGGGTTTCCAGATGAAGAAGTAT 2520  
2521 ctcttgagacaattggtcaaaattcaggctattaaaaatgagtgttcgattggtacttgaa 2580  
2521 CTCCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGTTTCGATGGCTACTTGGAA 2580  
2581 tgaataataatcagactaatcaggaactctcacttaagattgctcaacaacaatatgc 2640  
2581 TGAAAAATAATCACAGTAAATFCAGGAACCTTCTACTCTTAAAGATTGCTTAACAAATATTGC 2640  
2641 atagtatgagaccttacagaacaggggaaaattagttaaacagatatgtcacgtctga 2700  
2641 ATAGTATGAGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCAGCTCTGA 2700  
2701 gacttgcctggtgagtgctatttgaaactggtgacagaacccctgttaccatgaaatca 2760  
2701 GACTTGTCTGCTGGAGTGTCTATTGTGAAGCTGGCACAGAACCCTGTTTACCATGAATCA 2760  
2761 tcacattagaacaataatcagctatgtgcattagcta tcaacgatagaatgctatcaagtaa 2820  
2761 TCACATTAGAACAATATCAGCTATGTGCATTTAGCTATCAACAGATGAATGCTTATCAAGTAA 2820  
2821 gacaagtgttggcccaaaaacttcacaaagccctttcccgtttacgcttccacttaagt 2880  
2821 GACAAGTGTGTTGCCAGAAACTTCACAAAGGCCCTTTCCCGTTTACGGCTTCCACTTGTAGT 2880  
2881 atatggcaatctgctccctttgtcaaaagatcctgttaaaagagagaagagctcatgcta 2940  
2881 ATATGGCAATCTGTGCCCCCTTTGTCAAAAGATCCTGTAAAGGAGAGAAGACCTCATGCTA 2940  
2941 ggcgaattgttgtaaaaaataaattgtaagcggggagatctctgaagcagcatgcactg 3000  
2941 GGCAATGTTTGGTGA AAAATATAAATGTAAAGCGGAGTATCTGAAGCAGCATGCAGCTG 3000

3001 ttagtgaataattatgtctctctaccagagtgtgtgttctccatatacaattccacctt 3060  
3001 TTAGTGAAAAATTAATTGTCTCTCTTACCAGAGTATGTTGTTCATATACAATTCACCTTT 3060  
3061 tggcacatgacccagattatgtcaaaagtacaggatatgaacaacttaaaagtattaaag 3120  
3061 TGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACATTTAAAGATGTTAAAG 3120  
3121 aatgtctttggttgttctcggaataattaatgtgctaaaaatgaaaaataaacagtcacgctt 3180  
3121 AATGCTCTTGGTTGTCTTGGAAATATTAAATGGCTAAAAATGAAATAACAGTACACGCTT 3180  
3181 ttatcagaagaagtgttagaaaaatttaaaaaacaaagaatgcccagagaccagatgatg 3240  
3181 TTATCAGAAAGATGGTAGAAAATATTAAAGCAAAACAAAGATGCCCAAGGACCATGATG 3240  
3241 caaaaatgaatgaaaaaactgtcacactgtgtgtgatgttgccatgaatatcatcatgctcaa 3300  
3241 CAAAAATGAATGAAAAAAGTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAA 3300  
3301 agagtactacatacagtttggaaatctctctaaagaccgcggtactaccagctcgtttcttca 3360  
3301 AGAGTACTACATACAGTTTGGAAATCTCTTAAAGACCCCGGTACTACCAGCTCGTTCTTTCA 3360  
3361 ctcaacctgacaagaatttcagtaacacccaaaattatctcctcctcctgaaatgaatcat 3420  
3361 CTCAACCTGACACAAATTTTCACTAACCAAAAATTTATCTGCCCTCGTGAATGAATCAT 3420  
3421 ttttctcctctggaaaacctaataacacccaatgttcttaggagctgttaacaagccacttt 3480  
3421 TTTTCACTCTCTGGAAAAACCTTAAACCAACCAATGTTCTAGGAGCTGTTAAACAAGCCACTTT 3480  
3481 catcagcaggcaagcaatctcagaccocaaatcatcacgaatggaactgtgaagcaatgcaa 3540  
3481 CATCAGCAGGCAAGCAATCTCAGACCAAAATCATCACGAATGGAACACTGTAGGCAATGCAA 3540  
3541 qcacagctcaaaatcccaagctcctctggaagaataaaagggagcttgatagttcttgaaa 3600  
3541 GCACAGCTCAAAATCCAAAGCTCTCTGGAAGATAAAGGGGAGGCTTGATAGTTCTTGAAA 3600  
3601 tggatcagtgaaaaatgaagattacacaaatgtcttcaacctttccgggggaaaaaagtgt 3660  
3601 TGGATCAGTGAAAAATGAAGATTACAAATGTCTTCAACCTTTGCCGGGAAAAAGAGATG 3660  
3661 acaagagagacgactctgactcttgaaggtctgaattggaagaaccccttagagggcagaaaa 3720  
3661 ACAAGAGAGAGAGCTCTGATCTTGAAGGTCTGAATTGGAGAAGCCTAGAGGCAGAAAAA 3720  
3721 aaacgcccgtcacagaacagagagagaaattagttagtgatgacttgaactgaagttggtac 3780  
3721 AAACGCCGTGCACAGAACGGGAGGAGAAAATTTAGGTATGGATGACTTGACTAAGTTGGTAC 3780  
3781 aggaacgaacacctaaaggcagtcagcgaagtgcgaaaaagaggccatcacgcttccagaat 3840  
3781 AGGAACAGAGACCTTAAAGGCAGTCAGCGAAGTCGGAAGAAGAGGCCATACGCGCTCAGAAT 3840  
3841 ctgatgaacagcagtgccctgaggaagaagaggctcaaaagaagatatattagaanaatgaag 3900  
3841 CTGATGAACAGCAGTGCCCTGAGGAAAAAGAGGCTCAAAAGAAGATATATTAGAAAAATGAAG 3900  
3901 atgaacgaatagttccgccccaaaaagggtaaaagggcgcgaccaccaaaaacctcttggtg 3960  
3901 ATGAACAGATAGTCCCGCCAAAAAAGGGTAAAGAGGCCGACCAACCAAAACCTCTTGTGTG 3960  
3961 gaggtacacaaaaagaagagccaaatgaaaacttctaaaaaaggaagcaaaaaaaat 4020  
3961 GAGGTACACCAAAAAGAGAGGCCAACAAATGAAAACCTTCTTAAAAAAGGAAGCAAAAAAAT 4020  
4021 ctggaacctccagaccagagagagaggaagaagaagaagcaaaagtggaaataacggaac 4080  
4021 CTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAAGAAAGCAAAAGTGGAAATACGGAAC 4080



```
exon          3584...3689
/ gene="AS3"
/ number=31
exon          3690...4129
/ gene="AS3"
/ number=32
exon          4130...4354
/ gene="AS3"
/ number=33
exon          4355...4810
/ gene="AS3"
/ number=34

BASE COUNT   1755 a   944 c   1074 g   1480 t
ORIGIN

Query Match      80.3%; Score 4233; DB 97; Length 5253;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 5233; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 ccggagagccccggagtgagcggagtagcagtcgagcgcgcaacccggaggggtagaataattt 60
DB 1 CCGGAGAGCCCCGGAGTGAGCGAGTAGCGAGTCGGCAACCCGGAGGGGTAGAAATATTT 60
QY 61 ctgtcatggctcatccaagactagaccatagatggaataattacatctccgcctgggg 120
DB 61 CTGTCTATGGCTCATTCGAAGACTAGGACCAATGATGGAAAAATTACATATCCGCTGGGG 120
QY 121 tcaagaaatccagataaataatcctaagagagagatggtgagcagattaaagatgggtg 180
DB 121 TCAGGAAATATCAGATAAATATCTAAAGAGGAGATGTTGAGACGATTAAGATGGTTG 180
QY 181 tgaacacttttggataggaccaggagctctgaagaagaagaaggagctttttaaacc 240
DB 181 TGAACAACTTTTATGATATGGACAGGACTCTGAAGAAGAAAAAGGAGCTTTATTTAAACC 240
QY 241 tagcttacatcttgcctcaagattttttctcaagcactcctggtgaagaagtgcgttac 300
DB 241 TAGCTTACATCTTGCCTTCAGATATTTTCTCAAGCATCTCTGATAAAGATGTCGCTTAC 300
QY 301 tggtagctgccttgcctgatatttccagatttatgctcctgaagctcctcacat 360
DB 301 TGGTAGCTGCTGCCTTCGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACAT 360
QY 361 cccctgataaactaaaggatataatttatgtttatacaagacagttggaagggttagagg 420
DB 361 CCCCTGATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGCTAGAGG 420
QY 421 atacaagagcccaattcaataggtatttttatttacttgagaacattgcttgggtca 480
DB 421 ATACAAGAGCCCCACAGTCAATAGGTATTTTATTTGCTTGAGAACATTTGCTGGGTCA 480
QY 481 agtcataaactatgcttggatgagtagaagtagcaatgaaattttccaccagctataca 540
DB 481 AGTCATATAACATATGCTTTGAGTTAGAGATAGCAATGAATTTTCACCCAGCTATACA 540
QY 541 gaaccttatttcagttataaacaatggccacaatcagaagaatcccatatgacatggttag 600
DB 541 GAACTTATTTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGACATGGTAG 600
QY 601 acctatgagctctattatttgaagtgatacagtgctcctcaggagcttttggatacgg 660
DB 601 ACCTTATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGG 660
QY 661 ttttagtaacttggtacctgctcataagaatttaacaagaagcaagcatatgatttggcaa 720
DB 661 TTTTAGTAAATCTGCTACCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTTGGCAA 720
QY 721 agggcttactgaagaggacagctcaagctatttgagccatataattaccacttttttaatc 780
DB 721 AGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTATTAATC 780
QY 781 aggttctgatgctgggaaaaaacatctatcagcgatttgcagagcagatgcttctgacttaa 840
DB 781 AGGTTCTGATGCTGGGAAAAACATCTATCAGCGATTTGTGAGAGCATGCTCTGACTTAA 840
QY 841 ttttgagctctacaataattgatagtcatttctgctctctcttaccacagcttgagt 900
DB 841 TTTTGGAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCAGCTTGAAT 900
QY 901 ttaataaagagcaatgataatagaggagcgcctacaagtgtttaaactactggcaaaaa 960
DB 901 TTAATAAAGAGCAATGATGATGAGGAGCGCTACAAGTTGTTAAACTACTTGGCGAAAA 960
QY 961 tgtttggggcaaaagatttcagaattggcttctcaaaaaaagccacttttggcagtgctact 1020
DB 961 TGTTTGGGCAAAAGGATTCAGAAATTGGCTTCTCAAAAAAGCCACTTTGGCAGTGTCTACT 1020
QY 1021 tggcgaggttttaattgatattccatgtaccatccgcctggaatggtgaaatttgcagcc 1080
DB 1021 TGGCGAGGTTTAAATGATATCCATGTACCATCCGCTCGGAATGTGGAATTTGCTAGCC 1080
QY 1081 attgtctcatgaaccatctctgatttagcaaaaagacttaacagagtagtctttaaagtga 1140
DB 1081 ATTGTCTCATGAACCATCTCTGATTTAGCAAAAGACTTAAACAGAGTATCTTAAAGTGAGGT 1140
QY 1141 cacatgaccttgaggaaagctatttagacatgattgttatgtgcaatagttacagctgcta 1200
DB 1141 CACATGACCTTGAGGAAGCTATTAGACATGATGTTATGTTGTCATTAATGATGAGCTGCTA 1200
QY 1201 aaaaagatatcttctctggtgcaatgatcacttacttaatttgcagagagagaaacattag 1260
DB 1201 AAAAGGATATCTTCTGTCATGATCATTACTTAATTTTGTGGGAGAGAGAACATTTAG 1260
QY 1261 acaacgctggagagtagcaagaagccatgatgggacttgcacaaatttataagaat 1320
DB 1261 ACAACGATGGAGAGTAGCGAAAGAACCCATGATGGACTTGGCCAAATTTATAAGAAAT 1320
QY 1321 atgctttacagtcacagctggaaaagatgctgcgaaaaacagatagcatggtatcaagaaca 1380
DB 1321 ATGCTTTACAGTCACAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATGATCAAGACA 1380
QY 1381 aattgctacatatattatcaaaaatagattgtagatgcgactactgttgaaacgagatct 1440
DB 1381 AATTGCTACATATATATATCAAAATAGTATTGATGATGCTGACTTGTGTAACGGATCT 1440
QY 1441 ttgctcaatacagtggttctctcaaatattagaactacagaacgagtagaaatgcttata 1500
DB 1441 TTGCTCAATACATGCTTCTCTCAAAATTTAGAAACTACAGAACGGATGAATGCTTATAT 1500
QY 1501 actgtatgcacacatgaaatttaaatgctgtgaaagcattgaaatggtggaatgctc 1560
DB 1501 ACTGTATGCCACACTGGATTTAAATGCTGTGAAGCATTTGAATGAATGTGGAATGTC 1560
QY 1561 aaaaactgctccgacatcaagtaaaggatttgcctgacttgattgaagcaaccccaaacag 1620
DB 1561 AAAATCTGCTCCGACATCAAGTAAAGGATTTGCTGACTTGTATTAAGCAACCCAAAACAG 1620
QY 1621 atgcaagtgcgaagccatatttcaaaagtgtatggttattacaagaatttaccctgactc 1680
DB 1621 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGTATGTTATTACAGAGTGTACCTGATC 1680
QY 1681 ctggtaagctcagatttcatgaagaattcacacaggtgttagaagatgagaaaa 1740
DB 1681 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTCACACAGGTGTGTAGGAAGATGATGAGAAA 1740
QY 1741 taagaagcagttagaagtacttggtagtcccaacatgctccctgcagcagcaggtggaaggtt 1800
DB 1741 TAAGAAGCAGTTAGAGTACTTGTGTAGTCCAAACATGCTCCTGCAAGCAGGCTGAAGGTT 1800
QY 1801 gtgtgcgtgaataactaagaaggttggcacaaccccaaacagcactcaaatccttctctg 1860
DB 1801 GTGTGCGTGAATAACTAAGAAGTTGGGCAACCCCAACAGCCTCAAAATCCTTTCTCTGG 1860
QY 1861 aaatgatcaagttctcttggagagtagcagcttgcacatagataccgaatctatca 1920
DB 1861 AAATGATCAAGTTCTCTTGGAGAGTAGCAGCTTGCACATAGATACCgaatctatca 1920
```



TITLE Direct Submission  
JOURNAL Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University  
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA  
REFERENCE 4 (bases 1 to 5253)  
AUTHORS Geck, P., Szelele, J., Jimenez, J., Sonnenschein, C. and Soto, A. M.  
JOURNAL Direct Submission  
Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University  
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA  
REMARK Sequence update by submitter  
COMMENT On Apr 5, 1999 this sequence version replaced gi:4539617.  
FEATURES  
source  
1..5253  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13q12-q13"  
/cell\_line="LNCaP"  
/tissue="cancer cells"  
/note="exon 1 maps to p1 Artificial Chromosome PAC26H23;  
exons 2-24 map to cosmid 267p19; exons 25-34 map to p1  
Artificial Chromosome PAC 49010"  
1..5253  
/gene="AS3"  
1..46  
/gene="AS3"  
/number=1  
47..173  
/gene="AS3"  
/number=2  
66..4241  
/gene="AS3"  
/codon\_start=1  
/product="androgen-induced prostate proliferative shutoff  
associated protein"  
/protein\_id="AAD22134.2"  
/db\_xref="GI:4559410"  
/translation="WAHSTRNDGKITYPGVKTSKDEEMVRRRLKMYKTFMD  
MQDSEERELYNLAHLASDFELKHPDKDVRLLVACCLADIERLYAPEAPYSPDK  
LADIPFTRQLKGLDTSKSPQFNRYFYLENNIAWVKSYNICEFEEDSNEITQLYRT  
LFSVINNHQNVHMHVDMSSIICEGDTVSQELLDTVLNLPVAKNKNKOAYDLA  
KALLKRTAQAIIEPYITNFNQVLMGKTSISLSEHVSIDLILELYNIDSHLLSLPQ  
LEFKLSNDEERLQVLLAKMFGAKDSQKPLWQCYLGRFNDIHVIRLECV  
KEASHCLMNHDPDLKTEYLKVRSHDPEAIRHDVIVSIVTAARKDILLVHYQNSID  
VGERLDRKWRVKEAMGIAOIYKKYALQSAAGKDAAKOIAWIKDKLLHIYQNSID  
DRLLEVERIAQWVPHNLETTERMKCLYLYATLDLNAVKNEMWKNLKHQVKD  
LDDLKQPTDASVKAIFSVMVITKSLPPGKAQDFMKFTQVLEDDERKQLEVL  
VSPTCSKQAEQCVREITKKNPKOPTNPFLEMTFLIERIAPVHIDTESALIKQ  
VNSIDGTADDEGVPTDQATIRAGLELLKVLSTHPISFHSAETPESLACIKMDE  
KVAEAAOIFKNTGSKIEDPHIRSAALLPVLLHKSCKGPPQAKYAIHCIIAIFSSK  
CIVQVQVFAQLKHLKSLRLEPLRYMAICALKADPKVERRAHARQCLVKNNVREY  
LKQHAASVEKLLSLLEPYVVPYVTHLHADPDYVKVDIEQLDKVCELCWFLVLELMA  
KNENSHAFIRKVENIKOTKDAQGDDAKMEKLYTCDVAMNIIIMSKSTYISLESP  
KQPLVPARFTQPKNFNTKNYLPFGKMSFTPGPKTNYLVGAVNKLPSAGQSQ  
TKSSRMETVGNASSSNPSPGRIKGRDSSMDHSENEEDYTMSPLPKGRKDRDSS  
DLVRSLEKPRKPTVTEREELKGMDDTLKVDQRPKQSQRSRKRGHSTASEDEQ  
QWPEERKLEDILENEDEQSPKKGKGRPKPLGGGTPEEPTMKTSGKSKKSG  
PPAPEEEEEEQSGNTEQSKSKQHRVSRRAQQAESSESAIESTQSTPQKGRGP  
SKTPSPQPKKNV"  
174..377  
/gene="AS3"  
/number=3  
378..464  
/gene="AS3"  
/number=4  
465..562  
/gene="AS3"  
/number=5  
563..689  
/gene="AS3"  
/number=6  
690..770  
/gene="AS3"  
/number=7  
771..911  
/gene="AS3"  
/number=8  
912..1027  
/gene="AS3"  
/number=9  
1028..1122  
/gene="AS3"  
/number=10  
1123..1268  
/gene="AS3"  
/number=11  
1269..1420  
/gene="AS3"  
/number=12  
1421..1534  
/gene="AS3"  
/number=13  
1535..1616  
/gene="AS3"  
/number=14  
1617..1665  
/gene="AS3"  
/number=15  
1666..1805  
/gene="AS3"  
/number=16  
1806..1921  
/gene="AS3"  
/number=17  
1922..2027  
/gene="AS3"  
/number=18  
2028..2188  
/gene="AS3"  
/number=19  
2189..2312  
/gene="AS3"  
/number=20  
2313..2471  
/gene="AS3"  
/number=21  
2472..2540  
/gene="AS3"  
/number=22  
2541..2677  
/gene="AS3"  
/number=23  
2678..2801  
/gene="AS3"  
/number=24  
2802..3006  
/gene="AS3"  
/number=25  
3007..3121  
/gene="AS3"  
/number=26  
3122..3254  
/gene="AS3"  
/number=27  
3255..3374  
/gene="AS3"  
/number=28  
3375..3437  
/gene="AS3"  
/number=29  
3438..3583  
/gene="AS3"  
/number=30



885	27	0.5	69282	71	AC040912	Homo sapi	AC040912	Homo sapi	101041	83	AP002079	Homo sapi
886	27	0.5	69290	76	AC084206	Homo sapi	AC084206	Homo sapi	101077	68	AC024051	Homo sapi
887	27	0.5	69570	72	AC055831	Homo sapi	AC055831	Homo sapi	102064	85	AC002558	Homo sapi
888	27	0.5	70105	60	AC007732	Homo sapi	AC007732	Homo sapi	102092	78	AL138935	Homo sapi
889	27	0.5	70944	67	AC022723	Homo sapi	AC022723	Homo sapi	102714	91	AP000692	Homo sapi
890	27	0.5	71118	82	AP000586	Homo sapi	AP000586	Homo sapi	102892	90	AL359813	Human DNA
891	27	0.5	71983	80	AL443429	Human DNA	AL443429	Human DNA	103146	92	HS187N21	Human DNA
892	27	0.5	73029	78	AC090560	Homo sapi	AC090560	Homo sapi	103566	85	AC003087	Human DNA
893	27	0.5	73037	77	AC090247	Homo sapi	AC090247	Homo sapi	103694	85	AC004836	Homo sapi
894	27	0.5	73087	92	HS147M19	Homo sapi	HS147M19	Homo sapi	103739	84	HS047C9	Homo sapi
895	27	0.5	73607	90	AL162740	Human DNA	AL162740	Human DNA	103750	82	AL590433	Mus muscu
896	27	0.5	73701	64	AC016402	Homo sapi	AC016402	Homo sapi	104296	77	AC087771	Medicago
897	27	0.5	73754	82	AP000592	Homo sapi	AP000592	Homo sapi	104780	92	HS1185H19	Human DNA
898	27	0.5	74989	67	AC023333	Homo sapi	AC023333	Homo sapi	105407	93	AL332773	Human DNA
899	27	0.5	75582	89	AL138729	Human DNA	AL138729	Human DNA	105463	90	AL355272	Human DNA
900	27	0.5	75863	64	AC016059	Homo sapi	AC016059	Homo sapi	105576	92	HS106120B	Human DNA
901	27	0.5	76086	77	AC084246	Homo sapi	AC084246	Homo sapi	105643	86	AC005629	Homo sapi
902	27	0.5	77498	90	AL162505	Human DNA	AL162505	Human DNA	106168	87	AC011450	Homo sapi
903	27	0.5	78404	90	AL445198	Human DNA	AL445198	Human DNA	106508	85	AC005102	Homo sapi
904	27	0.5	78405	73	AC068094	Homo sapi	AC068094	Homo sapi	106664	94	AF130357	Mus muscu
905	27	0.5	78692	65	AC018370	Homo sapi	AC018370	Homo sapi	106827	85	AC004929	Homo sapi
906	27	0.5	80920	85	AC002036	Homo sapi	AC002036	Homo sapi	106835	79	AL157767	Homo sapi
907	27	0.5	81524	93	HS047C9	Homo sapi	HS047C9	Homo sapi	106921	85	AC004263	Homo sapi
908	27	0.5	81524	93	HS047C9	Homo sapi	HS047C9	Homo sapi	106988	85	AC001226	Genomic S
909	27	0.5	82359	85	AC004922	Homo sapi	AC004922	Homo sapi	107264	91	AP002078	Homo sapi
910	27	0.5	83059	91	AP001435	Homo sapi	AP001435	Homo sapi	107582	77	AC084697	Homo sapi
911	27	0.5	84300	93	HS104A17	Human DNA	HS104A17	Human DNA	107745	90	AL139042	Human DNA
912	27	0.5	84371	83	AP001963	Homo sapi	AP001963	Homo sapi	108063	67	AC022470	Homo sapi
913	27	0.5	84478	89	AL133462	Human DNA	AL133462	Human DNA	108329	90	AL353596	Human DNA
914	27	0.5	85565	89	AL136133	Human DNA	AL136133	Human DNA	109201	85	AC000403	Genomic S
915	27	0.5	85624	90	AL133933	Human DNA	AL133933	Human DNA	109672	60	AC008723	Homo sapi
916	27	0.5	85952	93	HS047C9	Human DNA	HS047C9	Human DNA	109694	14	F23A5	
917	27	0.5	86113	89	AF285442	Homo sapi	AF285442	Homo sapi	109878	90	AL390208	Human DNA
918	27	0.5	86137	77	AC090276	Homo sapi	AC090276	Homo sapi	110000	60	AC003656	5
919	27	0.5	86437	77	AC090276	Homo sapi	AC090276	Homo sapi	110000	60	AC008576	2
920	27	0.5	86437	77	AC090276	Homo sapi	AC090276	Homo sapi	110000	70	AC026388	1
921	27	0.5	86766	83	AP001872	Homo sapi	AP001872	Homo sapi	110000	76	AC079517	3
922	27	0.5	87022	82	AP001089	Homo sapi	AP001089	Homo sapi	110414	92	HS281H8	
923	27	0.5	87789	89	AL133270	Human DNA	AL133270	Human DNA	110885	92	HS191E19	Human DNA
924	27	0.5	87844	14	BNA245480	Brassica	BNA245480	Brassica	110913	82	AP000407	Homo sapi
925	27	0.5	88036	67	HUAC003119	Human Chr	AC021487	Homo sapi	110733	86	AC006375	Homo sapi
926	27	0.5	88309	66	AC021487	Homo sapi	AC021487	Homo sapi	111084	86	AC006486	Homo sapi
927	27	0.5	89015	60	AC008458	Human DNA	AL121509	Human DNA	111293	92	HS156L9	Homo sapi
928	27	0.5	89251	93	HS047C9	Human DNA	AP001252	Homo sapi				
929	27	0.5	89292	91	AP001252	Homo sapi	AC003062	Mouse Chr				
930	27	0.5	89743	94	AC003062	Mouse Chr	AC026450	Homo sapi				
931	27	0.5	90605	70	AC026450	Homo sapi	AF242432	Mus muscu				
932	27	0.5	90650	94	AF242432	Mus muscu	AC006186	Homo sapi				
933	27	0.5	91202	60	AC006186	Homo sapi	AL353688	Human DNA				
934	27	0.5	92569	90	AL353688	Human DNA	AC090117	Homo sapi				
935	27	0.5	93151	77	AC090117	Homo sapi	AC021825	Homo sapi				
936	27	0.5	93223	66	AC021825	Homo sapi	AP000246	Homo sapi				
937	27	0.5	94212	91	AP000246	Homo sapi	AC022465	Homo sapi				
938	27	0.5	95360	67	AC022465	Homo sapi	AL359694	Human DNA				
939	27	0.5	95369	90	AL359694	Human DNA	AC009894	Arabidops				
940	27	0.5	96489	12	AC009894	Arabidops	AP001243	Homo sapi				
941	27	0.5	96500	91	AP001243	Homo sapi	AL354890	Homo sapi				
942	27	0.5	97799	79	AL354890	Homo sapi	Z93930	Human DNA				
943	27	0.5	97835	92	HS292E10	Human DNA	AC005352	Homo sapi				
944	27	0.5	98543	85	AC005352	Homo sapi	AC008969	Homo sapi				
945	27	0.5	98562	87	AC008969	Homo sapi	AL096764	Human DNA				
946	27	0.5	99332	93	HS292E10	Human DNA	AL020990	Human DNA				
947	27	0.5	99342	92	HS292E10	Human DNA	AC006271	Homo sapi				
948	27	0.5	99344	86	AC006271	Homo sapi	AC005057	Homo sapi				
949	27	0.5	99370	85	AC005057	Homo sapi	AL512454	Human DNA				
950	27	0.5	99716	91	AL512454	Human DNA	AC006464	Homo sapi				
951	27	0.5	99908	86	AC006464	Homo sapi	AP000130	Homo sapi				
952	27	0.5	100000	91	AP000130	Homo sapi	AP000208	Homo sapi				
953	27	0.5	100039	88	AC002078	Homo sapi	AL019657	Human DNA				
954	27	0.5	100272	93	HS047C9	Human DNA	AC007359	Homo sapi				
955	27	0.5	100627	86	AC007359	Homo sapi	AC009489	Homo sapi				
956	27	0.5	100953	87	AC009489	Homo sapi	AC079392	Homo sapi				
957	27	0.5	101006	76	AC079392	Homo sapi						

## ALIGNMENTS

## RESULT 1

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

Human androgen-induced prostate proliferative shutoff associated

protein (AS3) mRNA, complete cds.

U95825

U95825.2

GI:4559409

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 5253)

Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Sonnenschein, C.

Androgen-induced proliferative shutoff in prostate cancer cells

Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)

2 (bases 1 to 5253)

Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Soto, A.M.

Early gene expression during androgen-induced inhibition of

proliferation of prostate cancer cells: a new suppressor candidate

on chromosome 13, in the BRCA2-Rb1 locus

J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)

99229875

3 (bases 1 to 5253)

Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Soto, A.M.



739	28	0.5	223398	78	AL137161	AL137161 Homo sapi	812	27	0.5	36767	85	AC004646	AC004646 Homo sapi
740	28	0.5	225300	77	AC084798	AC084798 Mus muscu	813	27	0.5	37170	93	HSE129H9	268224 Human DNA s
741	28	0.5	225371	81	AL445222	AL445222 Homo sapi	814	27	0.5	37655	86	AC007792	AC007792 Homo sapi
742	28	0.5	226841	84	HSAC002043	AC002043 Homo sapi	815	27	0.5	37930	86	AC005564	AC005564 Homo sapi
743	28	0.5	227008	82	AF001771	AP001771 Homo sapi	c 816	27	0.5	38258	85	AC000402	AC000402 Genomic s
744	28	0.5	229962	76	AC079520	AC079520 Mus muscu	c 817	27	0.5	38682	93	HSAC002077	AC002077 Human cos
745	28	0.5	237678	94	AC055818	AC055818 Mus muscu	c 818	27	0.5	39595	85	AC004509	AC004509 Homo sapi
746	28	0.5	249687	63	AC015693	AC015693 Homo sapi	c 819	27	0.5	39854	88	AC020954	AC020954 Homo sapi
747	28	0.5	252370	4	AE003529	AE003529 Drosophil	c 820	27	0.5	40582	91	AP000691	AP000691 Homo sapi
748	28	0.5	267685	68	AC024901	AC024901 Homo sapi	821	27	0.5	40839	88	AC026805	AC026805 Homo sapi
749	28	0.5	283512	80	AL365202	AL365202 Homo sapi	822	27	0.5	40709	97	L48038	L48038 Structure o
750	28	0.5	300000	91	AF002530	AP002530 Homo sapi	c 823	27	0.5	41015	92	HS590R14	AL049198 Human DNA
751	28	0.5	300691	60	AC005842	AC005842 Homo sapi	c 824	27	0.5	41052	85	AC002499	AC002499 Human Cos
752	28	0.5	321193	70	AC026340	AC026340 Homo sapi	c 825	27	0.5	41100	91	AP000540	AP000540 Homo sapi
753	28	0.5	336880	66	AC021052	AC021052 Homo sapi	c 826	27	0.5	41275	85	AC004182	AC004182 Homo sapi
754	28	0.5	340000	91	AP001724	AP001724 Homo sapi	c 827	27	0.5	41277	85	AC004180	AC004180 Homo sapi
755	27	0.5	213	14	SCB307346	AJ307346 Arabidops	c 828	27	0.5	41308	85	AC005203	AC005203 Homo sapi
756	27	0.5	213	15	TAE307350	AJ307350 Arabidops	c 829	27	0.5	41613	85	AC004790	AC004790 Homo sapi
757	27	0.5	306	88	AF103994	AF103994 Cebus ape	c 830	27	0.5	41696	86	AC005932	AC005932 Homo sapi
758	27	0.5	353	97	HUMCOL1A01	M32790 Human alpha	c 831	27	0.5	41766	90	AL359555	AL359555 Human DNA
759	27	0.5	359	97	HUMCOL1A1AM	M32798 Human alpha	c 832	27	0.5	42307	87	AC010510	AC010510 Homo sapi
760	27	0.5	390	13	ATH307357	AJ307357 Arabidops	c 833	27	0.5	42569	89	AL138915	AL138915 Human DNA
761	27	0.5	390	13	ATH307358	AJ307358 Arabidops	c 834	27	0.5	42854	91	AP001242	AP001242 Homo sapi
762	27	0.5	392	13	ATH307359	AJ307359 Arabidops	c 835	27	0.5	44078	85	AC002102	AC002102 Homo sapi
763	27	0.5	428	96	PWNSLRNB	L42376 Phytomonas	c 836	27	0.5	44145	87	AC010129	AC010129 Homo sapi
764	27	0.5	473	89	AF242445	AF242445 Homo sapi	837	27	0.5	44496	85	AC004760	AC004760 Homo sapi
765	27	0.5	473	94	MMU010984	AJ010984 Mus muscu	c 838	27	0.5	45186	91	AP001213	AP001213 Homo sapi
766	27	0.5	578	91	F32281S09	AF23289 Homo sapi	c 839	27	0.5	45333	92	HS10618A	AL049773 Human DNA
767	27	0.5	593	91	BC005936	BC005936 Homo sapi	c 840	27	0.5	46251	88	AD000671	AD000671 Homo sapi
768	27	0.5	667	94	AF150090	AF150090 Mus muscu	841	27	0.5	48014	77	AC087054	AC087054 Homo sapi
769	27	0.5	724	14	HVPSAG	X60158 H.vulgaris p	842	27	0.5	48716	77	AC087388	AC087388 Homo sapi
770	27	0.5	730	97	HSU67222	U67222 Human clone	843	27	0.5	51150	93	HSDJ913G4	AL045208 Human DNA
771	27	0.5	912	97	HSU67215	U67215 Human clone	844	27	0.5	51553	76	AC083779	AC083779 Homo sapi
772	27	0.5	1047	6	DD10DPGP1	M30467 D.discoideu	c 845	27	0.5	52286	77	AC090001	AC090001 Homo sapi
773	27	0.5	1055	94	BC002290	BC002290 Mus muscu	846	27	0.5	53038	78	AC090587	AC090587 Homo sapi
774	27	0.5	1115	96	SCU96099	U96099 Sarcophaga	847	27	0.5	54243	90	AL445215	AL445215 Human DNA
775	27	0.5	1136	9	A69662	A69662 Sequence 15	848	27	0.5	56250	77	AC090174	AC090174 Homo sapi
776	27	0.5	1144	85	AB045988	AB045988 Macaca fa	c 849	27	0.5	56835	69	AC025369	AC025369 Homo sapi
777	27	0.5	1467	92	HSN223951	AJ223951 Homo sapi	c 850	27	0.5	57084	90	AL158813	AL158813 Human DNA
778	27	0.5	1481	94	AF029212	AF029212 Rattus no	851	27	0.5	57304	85	AC004802	AC004802 Homo sapi
779	27	0.5	1598	6	DDMFOSLC	X54161 D.discoideu	852	27	0.5	58129	78	AC090404	AC090404 Homo sapi
780	27	0.5	1705	6	DDU20806	U20806 Dictyosteli	c 853	27	0.5	58129	78	AC090404	AC090404 Homo sapi
781	27	0.5	1770	97	LEHBB	M15734 Lemur (brow	854	27	0.5	60742	90	AL445208	AL445208 Human DNA
782	27	0.5	1835	13	AF281062	AF281062 Arabidops	855	27	0.5	61514	77	AC087697	AC087697 Homo sapi
783	27	0.5	2252	97	HUMCOL1A2	M55998 Human alpha	856	27	0.5	61544	77	AC090360	AC090360 Homo sapi
784	27	0.5	2351	88	AF068651	AF068651 Homo sapi	c 857	27	0.5	61795	78	AC090921	AC090921 Homo sapi
785	27	0.5	2454	97	HUMTSP	DI6517 Homo sapien	c 858	27	0.5	62232	63	AC015655	AC015655 Homo sapi
786	27	0.5	2618	14	CRCPRA	X69791 C.roseus cp	c 859	27	0.5	62275	76	AC079335	AC079335 Homo sapi
787	27	0.5	2785	94	AF162890S1	AF162890 Mus muscu	860	27	0.5	62366	71	AC036226	AC036226 Homo sapi
788	27	0.5	2848	94	BC004799	BC004799 Mus muscu	c 861	27	0.5	62964	76	AC080099	AC080099 Homo sapi
789	27	0.5	2917	94	AF225969	AF225969 Mus muscu	862	27	0.5	63437	93	HSJ964N17	AL050311 Human DNA
790	27	0.5	3447	9	A69663	A69663 Sequence 16	c 863	27	0.5	63773	12	AC074025	AC074025 Arabidops
791	27	0.5	3450	8	XL004707	U04707 Xenopus lae	864	27	0.5	64058	96	AC084113	AC084113 Homo sapi
792	27	0.5	5759	97	HUMFBNIGNA	L19896 Human fibril	865	27	0.5	64183	73	HSDJ429G5	AL078596 Human DNA
793	27	0.5	6070	83	AF085746	AF085746 Danio rer	c 866	27	0.5	64450	88	AC073893	AC073893 Homo sapi
794	27	0.5	6728	93	HSPALICO	Z74615 H.sapiens m	c 867	27	0.5	64814	14	NCB21J21	AL335929 Neurospor
795	27	0.5	6913	97	HSU19765	U19765 Human nucle	c 868	27	0.5	64898	78	AC090735	AC090735 Homo sapi
796	27	0.5	10234	94	AF094611	AF094611 Mus muscu	c 869	27	0.5	64988	78	AC090735	AC090735 Homo sapi
797	27	0.5	15302	90	AL163534	AL163534 Human DNA	c 870	27	0.5	65034	64	AC016067	AC016067 Homo sapi
798	27	0.5	15405	89	AF220542	AF220542 Homo sapi	871	27	0.5	65034	77	AC087290	AC087290 Homo sapi
799	27	0.5	16549	15	T31J12	AC006416 Arabidops	c 872	27	0.5	65034	77	AC087290	AC087290 Homo sapi
800	27	0.5	17906	94	AF071560	AF071560 Mus muscu	c 873	27	0.5	65427	78	AC091274	AC091274 Mus muscu
801	27	0.5	18609	88	AF071768	AF071768 Homo sapi	874	27	0.5	65472	78	AC091147	AC091147 Homo sapi
802	27	0.5	25483	91	AL450431	AL450431 Human DNA	875	27	0.5	65663	78	AC091305	AC091305 Homo sapi
803	27	0.5	28047	92	HS44083	AL022331 Human DNA	c 876	27	0.5	66227	70	AC026534	AC026534 Homo sapi
804	27	0.5	29122	91	AL513125	AL513125 Human DNA	c 877	27	0.5	66585	91	AL512378	AL512378 Human DNA
805	27	0.5	30758	89	AF278704	AF278704 Homo sapi	c 878	27	0.5	66975	89	AL135791	AL135791 Human DNA
806	27	0.5	32239	94	AB034241	AB034241 Rattus no	c 879	27	0.5	67150	69	AC025606	AC025606 Homo sapi
807	27	0.5	34164	78	AC090913	AC090913 Homo sapi	c 880	27	0.5	67722	77	AC087747	AC087747 Homo sapi
808	27	0.5	34544	6	CELY58GBA	AC006808 Caenorhab	881	27	0.5	68037	66	AC021757	AC021757 Homo sapi
809	27	0.5	35687	85	AC004779	AC004779 Homo sapi	882	27	0.5	68279	77	AC090165	AC090165 Homo sapi
810	27	0.5	35882	97	HSU41384	U41384 Human small	883	27	0.5	68407	72	AC048372	AC048372 Homo sapi
811	27	0.5	36255	64	AC016208	AC016208 Homo sapi	884	27	0.5	68763	72	AC090718	AC090718 Homo sapi



593	28	0.5	179473	64	AC016310	AC016310	Homo sapi	666	28	0.5	190709	60	AC008435	AC008435	Homo sapi
594	28	0.5	179485	86	AC005939	AC005939	Homo sapi	c 667	28	0.5	190821	86	AC006111	AC006111	Homo sapi
595	28	0.5	179597	73	AC068447	AC068447	Homo sapi	c 668	28	0.5	190907	85	AC005261	AC005261	Homo sapi
596	28	0.5	179613	81	AL390881	AL390881	Homo sapi	c 669	28	0.5	191422	67	AC022318	AC022318	Homo sapi
597	28	0.5	179704	70	AC026627	AC026627	Homo sapi	c 670	28	0.5	191877	86	AC007207	AC007207	Homo sapi
598	28	0.5	179797	75	AC073865	AC073865	Homo sapi	671	28	0.5	191904	86	AC005699	AC005699	Homo sapi
599	28	0.5	179929	86	AC006270	AC006270	Homo sapi	c 672	28	0.5	192126	91	DJ270M14	AC074021	Homo sapi
600	28	0.5	179955	67	AC023247	AC023247	Homo sapi	c 673	28	0.5	192318	75	AC074021	AC074021	Homo sapi
601	28	0.5	180038	73	AC068393	AC068393	Homo sapi	674	28	0.5	192610	79	AL355341	AL355341	Homo sapi
602	28	0.5	180129	75	AC074173	AC074173	Mus muscu	675	28	0.5	192782	75	AC079276	AC079276	Mus muscu
603	28	0.5	180254	76	AC084001	AC084001	Homo sapi	676	28	0.5	193020	89	AL138759	AL138759	Human DNA
604	28	0.5	180397	86	AC007035	AC007035	Homo sapi	677	28	0.5	193086	78	AF286887	AF286887	Homo sapi
605	28	0.5	180435	72	AC060812	AC060812	Homo sapi	678	28	0.5	193513	75	AC074372	AC074372	Homo sapi
606	28	0.5	180596	87	AC009570	AC009570	Homo sapi	c 679	28	0.5	193573	78	AL138921	AL138921	Homo sapi
607	28	0.5	180686	94	AF267747	AF267747	Mus muscu	c 680	28	0.5	194032	65	AC019150	AC019150	Homo sapi
608	28	0.5	180791	68	AC023977	AC023977	Homo sapi	681	28	0.5	194036	81	AL513479	AL513479	Homo sapi
609	28	0.5	180907	72	AC048383	AC048383	Homo sapi	c 682	28	0.5	194497	88	AC021887	AC021887	Homo sapi
610	28	0.5	181485	68	AC024036	AC024036	Homo sapi	c 683	28	0.5	194612	90	AL357354	AL357354	Human DNA
611	28	0.5	181487	78	AL139410	AL139410	Homo sapi	c 684	28	0.5	194794	62	AC012371	AC012371	Homo sapi
612	28	0.5	181823	71	AC040167	AC040167	Homo sapi	685	28	0.5	194935	69	AC026113	AC026113	Homo sapi
613	28	0.5	181930	78	AC091297	AC091297	Sus scrofa	686	28	0.5	194942	77	AC087884	AC087884	Homo sapi
614	28	0.5	182158	81	AL391116	AL391116	Homo sapi	687	28	0.5	195470	76	AC079367	AC079367	Mus muscu
615	28	0.5	182189	86	AC007333	AC007333	Homo sapi	688	28	0.5	195529	69	AC025612	AC025612	Homo sapi
616	28	0.5	182262	68	AC023963	AC023963	Homo sapi	689	28	0.5	196379	78	AL139319	AL139319	Homo sapi
617	28	0.5	182303	60	AC007638	AC007638	Homo sapi	690	28	0.5	196751	69	AC025494	AC025494	Homo sapi
618	28	0.5	182799	74	AC073526	AC073526	Homo sapi	c 691	28	0.5	196986	82	AF000941	AF000941	Homo sapi
619	28	0.5	182919	73	AC068779	AC068779	Homo sapi	c 692	28	0.5	197054	82	AL590036	AL590036	Homo sapi
620	28	0.5	183029	67	AC022067	AC022067	Homo sapi	c 693	28	0.5	197189	65	AC019168	AC019168	Homo sapi
621	28	0.5	183490	71	AC034244	AC034244	Homo sapi	c 694	28	0.5	197523	65	AC018883	AC018883	Homo sapi
622	28	0.5	183550	64	AC015913	AC015913	Homo sapi	c 695	28	0.5	197692	61	AC009554	AC009554	Homo sapi
623	28	0.5	183680	60	AC009077	AC009077	Homo sapi	696	28	0.5	198333	81	AL445673	AL445673	Homo sapi
624	28	0.5	183692	81	AL451125	AL451125	Homo sapi	697	28	0.5	198388	74	AC073426	AC073426	Homo sapi
625	28	0.5	183916	89	AF111169	AF111169	Homo sapi	c 698	28	0.5	199243	71	AC040925	AC040925	Homo sapi
626	28	0.5	184012	87	AC017080	AC017080	Homo sapi	c 699	28	0.5	200490	69	AC025548	AC025548	Homo sapi
627	28	0.5	184137	70	AC026472	AC026472	Homo sapi	c 700	28	0.5	200588	74	AC073266	AC073266	Homo sapi
628	28	0.5	184148	77	AC084413	AC084413	Mus muscu	701	28	0.5	200891	91	AL451146	AL451146	Human DNA
629	28	0.5	184181	90	AL158141	AL158141	Human DNA	702	28	0.5	201143	92	HS150C2	AL022318	Human DNA
630	28	0.5	184655	76	AC079379	AC079379	Homo sapi	703	28	0.5	201290	77	AC084375	AC084375	Homo sapi
631	28	0.5	184834	88	AC026122	AC026122	Homo sapi	c 704	28	0.5	201685	79	AL353743	AL353743	Homo sapi
632	28	0.5	184840	66	AC021422	AC021422	Homo sapi	c 705	28	0.5	202267	91	DJ293M10	AF111167	Homo sapi
633	28	0.5	184852	79	AL162732	AL162732	Homo sapi	c 706	28	0.5	202596	65	AC018441	AC018441	Homo sapi
634	28	0.5	184864	63	AC013553	AC013553	Homo sapi	c 707	28	0.5	202593	72	AC063955	AC063955	Homo sapi
635	28	0.5	185019	66	AC021231	AC021231	Homo sapi	c 708	28	0.5	202719	86	AC006379	AC006379	Homo sapi
636	28	0.5	185143	69	AC025861	AC025861	Homo sapi	c 709	28	0.5	202998	72	AC062014	AC062014	Homo sapi
637	28	0.5	185173	60	AC009056	AC009056	Homo sapi	c 710	28	0.5	203046	80	AC006342	AC006342	Homo sapi
638	28	0.5	185186	79	AL158826	AL158826	Homo sapi	711	28	0.5	203274	80	AL356957	AL356957	Homo sapi
639	28	0.5	185203	79	AL161649	AL161649	Homo sapi	712	28	0.5	204201	81	AL589666	AL589666	Homo sapi
640	28	0.5	185207	77	AC084350	AC084350	Homo sapi	c 713	28	0.5	204267	65	AC019071	AC019071	Homo sapi
641	28	0.5	185331	62	AC012059	AC012059	Homo sapi	714	28	0.5	204809	66	AC021133	AC021133	Homo sapi
642	28	0.5	185866	72	AC048389	AC048389	Homo sapi	c 715	28	0.5	205307	61	AC009968	AC009968	Homo sapi
643	28	0.5	185688	65	AC018805	AC018805	Homo sapi	c 716	28	0.5	205430	83	AF002756	AF002756	Homo sapi
644	28	0.5	185764	69	AC025289	AC025289	Homo sapi	717	28	0.5	205499	68	AC023352	AC023352	Homo sapi
645	28	0.5	185820	91	HS1009E24	HS1009E24	Human DNA	718	28	0.5	205993	62	AC011952	AC011952	Homo sapi
646	28	0.5	185960	60	AC008245	AC008245	Homo sapi	719	28	0.5	206578	70	AC026529	AC026529	Homo sapi
647	28	0.5	185994	60	AC002042	AC002042	Homo sapi	720	28	0.5	206647	91	AF002898	AF002898	Homo sapi
648	28	0.5	186221	62	AC012530	AC012530	Homo sapi	c 721	28	0.5	206848	72	AC063965	AC063965	Homo sapi
649	28	0.5	186321	86	AC006511	AC006511	Homo sapi	c 722	28	0.5	207170	83	AF003484	AF003484	Homo sapi
650	28	0.5	187386	69	AC025554	AC025554	Homo sapi	c 723	28	0.5	208145	72	AC053504	AC053504	Homo sapi
651	28	0.5	187451	74	AC073216	AC073216	Homo sapi	724	28	0.5	208763	79	AL353791	AL353791	Homo sapi
652	28	0.5	187647	74	AC073220	AC073220	Homo sapi	c 725	28	0.5	209303	66	AC020690	AC020690	Homo sapi
653	28	0.5	187877	67	AC023251	AC023251	Homo sapi	726	28	0.5	209845	70	AC026401	AC026401	Homo sapi
654	28	0.5	187955	73	AC067759	AC067759	Homo sapi	c 727	28	0.5	210946	79	AL353620	AL353620	Homo sapi
655	28	0.5	188152	69	AC025847	AC025847	Homo sapi	c 728	28	0.5	211442	64	AC016886	AC016886	Homo sapi
656	28	0.5	188603	76	AC079790	AC079790	Homo sapi	729	28	0.5	216288	80	AL390116	AL390116	Homo sapi
657	28	0.5	188684	90	AL355312	AL355312	Human DNA	730	28	0.5	216387	91	DJ534K4	AF109907	Homo sapi
658	28	0.5	189067	64	AC016810	AC016810	Homo sapi	731	28	0.5	216842	63	AC013726	AC013726	Homo sapi
659	28	0.5	189287	73	AC068362	AC068362	Homo sapi	c 732	28	0.5	216842	63	AC013726	AC013726	Homo sapi
660	28	0.5	189592	65	AC019146	AC019146	Homo sapi	c 733	28	0.5	218836	87	AC016734	AC016734	Homo sapi
661	28	0.5	189730	76	AC084044	AC084044	Mus muscu	c 734	28	0.5	219139	77	AC084418	AC084418	Homo sapi
662	28	0.5	189835	66	AC021011	AC021011	Homo sapi	735	28	0.5	219820	87	AC008738	AC008738	Homo sapi
663	28	0.5	189953	90	AL445435	AL445435	Human DNA	c 736	28	0.5	219935	85	AC005015	AC005015	Homo sapi
664	28	0.5	190000	60	AC004479	AC004479	Homo sapi	737	28	0.5	220553	82	AP001025	AP001025	Homo sapi
665	28	0.5	190440	90	AL357515	AL357515	Human DNA	c 738	28	0.5	221484	65	AC019095	AC019095	Homo sapi



C 447	28	0.5	158666	67	AC022394	Homo sapi	520
C 448	28	0.5	158760	69	AC026098	Homo sapi	C 521
C 449	28	0.5	159096	90	AL161781	Human DNA	C 522
C 450	28	0.5	159175	74	AC069421	Homo sapi	C 523
C 451	28	0.5	159508	65	AC018573	Homo sapi	C 524
C 452	28	0.5	159705	79	AL162613	Homo sapi	C 525
C 453	28	0.5	159859	75	AC074010	Homo sapi	C 526
C 454	28	0.5	160109	90	AL161727	Human DNA	C 527
C 455	28	0.5	160169	64	AC016660	Homo sapi	C 528
C 456	28	0.5	160193	69	AC025605	Homo sapi	C 529
C 457	28	0.5	160383	63	AC015523	Homo sapi	C 530
C 458	28	0.5	160590	61	AC009663	Homo sapi	C 531
C 459	28	0.5	160644	88	AC068770	Homo sapi	C 532
C 460	28	0.5	160696	88	AC090937	Homo sapi	C 533
C 461	28	0.5	160722	64	AC016963	Homo sapi	C 534
C 462	28	0.5	160759	83	AP002376	Homo sapi	C 535
C 463	28	0.5	160831	69	AC025015	Homo sapi	C 536
C 464	28	0.5	160892	81	AL512410	Homo sapi	C 537
C 465	28	0.5	160956	63	AC013637	Homo sapi	C 538
C 466	28	0.5	161155	88	AF017104	Homo sapi	C 539
C 467	28	0.5	161366	76	AC079848	Homo sapi	C 540
C 468	28	0.5	161514	88	AC078845	Homo sapi	C 541
C 469	28	0.5	161810	60	AC009082	Homo sapi	C 542
C 470	28	0.5	161979	81	AL451081	Homo sapi	C 543
C 471	28	0.5	162075	92	HS127D3	Homo sapi	C 544
C 472	28	0.5	162081	86	AC007542	Homo sapi	C 545
C 473	28	0.5	162105	70	AC026483	Homo sapi	C 546
C 474	28	0.5	162416	88	AC026882	Homo sapi	C 547
C 475	28	0.5	162426	82	AL590377	Homo sapi	C 548
C 476	28	0.5	162602	74	AC073633	Homo sapi	C 549
C 477	28	0.5	162617	86	AC007501	Homo sapi	C 550
C 478	28	0.5	162797	67	AC023219	Homo sapi	C 551
C 479	28	0.5	162829	81	AL513185	Homo sapi	C 552
C 480	28	0.5	163567	80	AL365325	Homo sapi	C 553
C 481	28	0.5	163608	69	AC026081	Homo sapi	C 554
C 482	28	0.5	164039	74	AC069354	Homo sapi	C 555
C 483	28	0.5	164247	79	AL158082	Homo sapi	C 556
C 484	28	0.5	164347	71	AC034124	Homo sapi	C 557
C 485	28	0.5	164352	68	AC024008	Homo sapi	C 558
C 486	28	0.5	164463	80	AL365310	Homo sapi	C 559
C 487	28	0.5	164494	87	AC020898	Homo sapi	C 560
C 488	28	0.5	164757	81	AL391215	Homo sapi	C 561
C 489	28	0.5	164823	66	AC009567	Homo sapi	C 562
C 490	28	0.5	165255	66	AC021421	Homo sapi	C 563
C 491	28	0.5	165287	77	AC087742	Homo sapi	C 564
C 492	28	0.5	165380	78	AC090529	Homo sapi	C 565
C 493	28	0.5	165537	73	AC068884	Homo sapi	C 566
C 494	28	0.5	166154	69	AC025779	Homo sapi	C 567
C 495	28	0.5	166549	69	AC025990	Homo sapi	C 568
C 496	28	0.5	166582	77	AC084398	Homo sapi	C 569
C 497	28	0.5	166797	80	AL359094	Homo sapi	C 570
C 498	28	0.5	167327	80	AL355979	Homo sapi	C 571
C 499	28	0.5	167603	73	AC068261	Homo sapi	C 572
C 500	28	0.5	167712	68	AC023809	Mus muscu	C 573
C 501	28	0.5	168111	69	AC025788	Homo sapi	C 574
C 502	28	0.5	168210	87	AC018719	Homo sapi	C 575
C 503	28	0.5	168347	69	AC025336	Homo sapi	C 576
C 504	28	0.5	168621	72	AC046173	Homo sapi	C 577
C 505	28	0.5	168665	72	AC053491	Homo sapi	C 578
C 506	28	0.5	168749	60	AC008573	Homo sapi	C 579
C 507	28	0.5	169162	62	AC012479	Homo sapi	C 580
C 508	28	0.5	169504	91	CNS010DTV	Human chr	C 581
C 509	28	0.5	169524	67	AC023111	Homo sapi	C 582
C 510	28	0.5	169622	80	AL359761	Homo sapi	C 583
C 511	28	0.5	170278	66	AC020586	Homo sapi	C 584
C 512	28	0.5	170278	66	AC020586	Homo sapi	C 585
C 513	28	0.5	170399	93	HS0J680D5	Human DNA	C 586
C 514	28	0.5	170420	62	AC011612	Homo sapi	C 587
C 515	28	0.5	170609	77	AC084863	Papio cyn	C 588
C 516	28	0.5	170623	90	AL391375	Human DNA	C 589
C 517	28	0.5	170923	64	AC016181	Homo sapi	C 590
C 518	28	0.5	171115	65	AC018846	Homo sapi	C 591
C 519	28	0.5	171135	81	AL441886	Homo sapi	C 592
AL354669	Homo sapi	79	AL354669				
AC021933	Homo sapi	67	AC021933				
AC010941	Homo sapi	61	AC010941				
AC067956	Homo sapi	88	AC067956				
AL356152	Homo sapi	80	AL356152				
AC021994	Homo sapi	67	AC021994				
AC089988	Homo sapi	77	AC089988				
AC020716	Homo sapi	87	AC020716				
AP000846	Homo sapi	82	AP000846				
AC025702	Homo sapi	69	AC025702				
AC018498	Homo sapi	65	AC018498				
AC067835	Homo sapi	73	AC067835				
AL442004	Homo sapi	81	AL442004				
AC006337	Homo sapi	86	AC006337				
AC017061	Homo sapi	64	AC017061				
AC008652	Homo sapi	60	AC008652				
AC026934	Homo sapi	70	AC026934				
AL359472	Homo sapi	80	AL359472				
AC018889	Homo sapi	63	AC018889				
AC015720	Homo sapi	65	AC015720				
AC012442	Homo sapi	87	AC012442				
AL137001	Homo sapi	78	AL137001				
AL355103	Human chr	91	CNS05TCM				
AC032044	Homo sapi	71	AC032044				
AC018999	Homo sapi	65	AC018999				
AL451062	Homo sapi	81	AL451062				
AC063950	Homo sapi	72	AC063950				
AC009127	Homo sapi	61	AC009127				
AP003124	Homo sapi	83	AP003124				
AC016106	Homo sapi	64	AC016106				
AL353718	Homo sapi	79	AL353718				
AP001840	Homo sapi	82	AP001840				
AC009087	Homo sapi	87	AC009087				
AP001363	Homo sapi	82	AP001363				
AL354659	Homo sapi	79	AL354659				
AC021490	Homo sapi	66	AC021490				
AC079897	Homo sapi	76	AC079897				
AC020621	Homo sapi	66	AC020621				
AL135783	Human DNA	89	AL135783				
AC019144	Homo sapi	65	AC019144				
AC025273	Homo sapi	69	AC025273				
AC025517	Homo sapi	69	AC025517				
AL353600	Homo sapi	79	AL353600				
AC010809	Homo sapi	61	AC010809				
AL391561	Homo sapi	81	AL391561				
AC013715	Homo sapi	63	AC013715				
AC009867	Homo sapi	61	AC009867				
AL138891	Homo sapi	78	AL138891				
AL079352	Human chr	91	CNS00M8V				
AC004887	Homo sapi	85	AC004887				
AC067925	Homo sapi	73	AC067925				
AC009967	Homo sapi	61	AC009967				
AC055732	Homo sapi	72	AC055732				
AC012036	Homo sapi	86	AC012036				
AC005409	Homo sapi	75	AC005409				
AC079019	Homo sapi	75	AC079019				
AC067739	Homo sapi	73	AC067739				
AC026384	Mus muscu	94	AC026384				
AC012322	Homo sapi	62	AC012322				
AC005826	Homo sapi	86	AC005826				
AC060829	Homo sapi	72	AC060829				
AC011737	Homo sapi	62	AC011737				
AC024446	Homo sapi	68	AC024446				
AL121748	Human DNA	93	HS342D11				
AC025341	Homo sapi	69	AC025341				
AC027085	Homo sapi	70	AC027085				
AL353807	Human DNA	64	AL353807				
AL353578	Homo sapi	90	AL353578				
AC068891	Homo sapi	79	AC068891				
AC002492	Human chr	73	AC002492				
AL356596	Homo sapi	97	HUAC002492				
AP003476	Homo sapi	83	CNS05TDF				



C 301	28	0.5	66671	78	AC090795	Homo sapi	AC090795	C 374	28	0.5	138041	91	CNS01DU1	AL133245	BAC sequ
C 302	28	0.5	70184	87	AC012441	Homo sapi	AC012441	375	28	0.5	138532	85	AC004812	Homo sapi	
C 303	28	0.5	72254	90	AL391646	Human DNA	AL391646	376	28	0.5	138613	77	AC084702	Homo sapi	
C 304	28	0.5	72807	68	AC023525	Homo sapi	AC023525	C 377	28	0.5	138814	73	AC068997	Homo sapi	
C 305	28	0.5	73403	79	AL355505	Homo sapi	AL355505	C 378	28	0.5	138839	97	HUAC002400	Human Chr	
C 306	28	0.5	74961	79	AL162258	Homo sapi	AL162258	C 379	28	0.5	139485	69	AC025597	Homo sapi	
C 307	28	0.5	75110	74	AC069115	Homo sapi	AC069115	380	28	0.5	139627	90	AL357117	Human DNA	
C 308	28	0.5	76241	93	HS0336M4	Human DNA	AL096775	381	28	0.5	139904	92	HS97D16	Homo sapi	
C 309	28	0.5	76660	69	AC025153	Homo sapi	AC025153	382	28	0.5	139973	67	AC022650	Homo sapi	
C 310	28	0.5	79612	85	AC004073	Human Chr	AC004073	C 383	28	0.5	140738	89	AF129408	Homo sapi	
C 311	28	0.5	81854	86	AC006396	Homo sapi	AC006396	384	28	0.5	141277	90	AL357045	Human DNA	
C 312	28	0.5	82001	12	AB005242	Arabidops	AB005242	385	28	0.5	141514	80	AL359095	Homo sapi	
C 313	28	0.5	82404	89	AL133399	Human DNA	AL133399	C 386	28	0.5	141614	82	AL590244	Homo sapi	
C 314	28	0.5	84122	89	AL133463	Human DNA	AL133463	387	28	0.5	142908	81	AL513172	Homo sapi	
C 315	28	0.5	86064	81	AL391554	Homo sapi	AL391554	388	28	0.5	143183	70	AC026927	Homo sapi	
C 316	28	0.5	86519	91	AP001860	Homo sapi	AP001860	389	28	0.5	143200	88	AC022081	Homo sapi	
C 317	28	0.5	87039	12	AC016529	Arabidops	AC016529	390	28	0.5	143369	86	AC006316	Homo sapi	
C 318	28	0.5	89820	86	AC007868	Genomic S	AC007868	391	28	0.5	143754	80	AL359093	Homo sapi	
C 319	28	0.5	92969	93	HS7687F11	Human DNA	AL117334	C 392	28	0.5	143754	80	AL359093	Homo sapi	
C 320	28	0.5	95565	85	AC000379	Human Chr	AC000379	C 393	28	0.5	144114	79	AL161453	Homo sapi	
C 321	28	0.5	96202	89	AL138824	Human DNA	AL138824	394	28	0.5	144168	66	AC021203	Homo sapi	
C 322	28	0.5	96726	84	HS1178F13	Homo sapi	AL118500	C 395	28	0.5	144660	60	AC008650	Homo sapi	
C 323	28	0.5	97982	91	AP000687	Homo sapi	AP000687	C 396	28	0.5	144967	60	AC008963	Homo sapi	
C 324	28	0.5	98869	62	AC011496	Homo sapi	AC011496	397	28	0.5	145253	86	AC006329	Homo sapi	
C 325	28	0.5	98933	88	AC025435	Homo sapi	AC025435	398	28	0.5	146093	78	AL139106	Homo sapi	
C 326	28	0.5	99011	60	AC004937	Homo sapi	AC004937	C 399	28	0.5	146170	75	AC079224	Homo sapi	
C 327	28	0.5	100301	92	HS317E23	Human DNA	AL020996	C 400	28	0.5	146260	60	AC008769	Homo sapi	
C 328	28	0.5	100506	83	AP002020	Homo sapi	AP002020	401	28	0.5	146372	90	AL359174	Human DNA	
C 329	28	0.5	101370	85	AC003962	Homo sapi	AC003962	402	28	0.5	146385	87	AC010534	Homo sapi	
C 330	28	0.5	102910	85	AB041512	Homo sapi	AB041512	C 403	28	0.5	147094	62	AC012027	Homo sapi	
C 331	28	0.5	104025	89	AL138679	Human DNA	AL138679	C 404	28	0.5	147786	61	AC010299	Homo sapi	
C 332	28	0.5	104147	85	AC004703	Homo sapi	AC004703	C 405	28	0.5	147945	62	AC012244	Homo sapi	
C 333	28	0.5	104269	70	AC026448	Homo sapi	AC026448	406	28	0.5	148515	83	CNS01RKG	Homo sapi	
C 334	28	0.5	106830	63	AC013394	Homo sapi	AC013394	407	28	0.5	148641	92	HS1119D9	Human DNA	
C 335	28	0.5	107260	90	AL139352	Human DNA	AL139352	C 408	28	0.5	149087	90	AL157826	Human DNA	
C 336	28	0.5	107418	92	HS551E13	Human DNA	AL022163	C 409	28	0.5	149253	62	AC012245	Homo sapi	
C 337	28	0.5	107429	93	HS3214M20	Human DNA	AL121969	410	28	0.5	149411	65	AC019029	Homo sapi	
C 338	28	0.5	108090	87	AC009533	Homo sapi	AC009533	411	28	0.5	149480	97	HU095740	Human Chrom	
C 339	28	0.5	108400	85	AC005193	Homo sapi	AC005193	412	28	0.5	149605	86	AC005522	Homo sapi	
C 340	28	0.5	109524	93	HSJ305B16	Human DNA	AL047972	413	28	0.5	150765	68	AC024326	Homo sapi	
C 341	28	0.5	110000	62	AC011600.2	Continuation (3 of	AL138724	C 414	28	0.5	150846	89	AL138724	Human DNA	
C 342	28	0.5	110000	64	AC015817.0	Homo sapi	AC015817	415	28	0.5	151851	84	CNS01RHL	Homo sapi	
C 343	28	0.5	110000	85	AC005079.0	Homo sapi	AC005079	C 416	28	0.5	151933	74	AC073199	Homo sapi	
C 344	28	0.5	110384	85	AC000117	Homo sapi	AC000117	C 417	28	0.5	152058	69	AC025232	Homo sapi	
C 345	28	0.5	111035	88	AC022405	Homo sapi	AC022405	418	28	0.5	152514	66	AC021329	Homo sapi	
C 346	28	0.5	111868	64	AC015848	Homo sapi	AC015848	C 419	28	0.5	152612	72	AC053515	Homo sapi	
C 347	28	0.5	111977	82	AP000717	Homo sapi	AP000717	C 420	28	0.5	152798	85	AC004057	Homo sapi	
C 348	28	0.5	112705	83	AP002025	Homo sapi	AP002025	C 421	28	0.5	153225	63	AC013581	Homo sapi	
C 349	28	0.5	112769	93	HS7635E18	Human DNA	AL109811	422	28	0.5	153493	69	AC026215	Homo sapi	
C 350	28	0.5	114317	22	AC012017	Human DNA	AC012017	C 423	28	0.5	154788	92	HS325F22	Human DNA	
C 351	28	0.5	115841	15	U95973	Arabidops	U95973	424	28	0.5	154848	63	AC012616	Homo sapi	
C 352	28	0.5	124474	90	AL158012	Human DNA	AL158012	C 425	28	0.5	154867	7	SSC131112	Sus scrofa	
C 353	28	0.5	125202	85	AC004858	Homo sapi	AC004858	426	28	0.5	155112	80	AL357521	Homo sapi	
C 354	28	0.5	127178	85	AC005160	Homo sapi	AC005160	427	28	0.5	155493	87	AC008825	Homo sapi	
C 355	28	0.5	127606	80	AL356298	Homo sapi	AL356298	C 428	28	0.5	155745	67	AC021941	Homo sapi	
C 356	28	0.5	127867	92	HS694B14	Human DNA	AL031673	C 429	28	0.5	155868	79	AL353779	Homo sapi	
C 357	28	0.5	129491	78	AL138842	Homo sapi	AL138842	430	28	0.5	156159	67	AC023325	Homo sapi	
C 358	28	0.5	131159	82	AP000743	Homo sapi	AP000743	C 431	28	0.5	156270	76	AC083951	Homo sapi	
C 359	28	0.5	131245	89	AL133320	Human DNA	AL133320	432	28	0.5	156372	78	AL139138	Homo sapi	
C 360	28	0.5	131587	80	AL358075	Homo sapi	AL358075	433	28	0.5	157248	85	AC002312	Human Chr	
C 361	28	0.5	132275	72	AC048388	Homo sapi	AC048388	434	28	0.5	157302	81	AL445209	Homo sapi	
C 362	28	0.5	132353	60	AC002355	Homo sapi	AC002355	435	28	0.5	157419	69	AC025389	Homo sapi	
C 363	28	0.5	132805	92	HS339A18	Human DNA	U97054	C 436	28	0.5	157666	86	AC006989	Homo sapi	
C 364	28	0.5	135734	60	AC002346	Homo sapi	AC002346	C 437	28	0.5	157814	71	AC037460	Homo sapi	
C 365	28	0.5	135805	74	AC069511	Homo sapi	AC069511	438	28	0.5	157871	87	AC012099	Homo sapi	
C 366	28	0.5	135961	66	AC021288	Homo sapi	AC021288	C 439	28	0.5	157930	81	AL391065	Homo sapi	
C 367	28	0.5	136124	92	HS394P21	Homo sapi	AL021528	440	28	0.5	157949	90	AL355578	Human DNA	
C 368	28	0.5	136385	86	AC006006	Homo sapi	AC006006	C 441	28	0.5	158116	79	AL353627	Homo sapi	
C 369	28	0.5	136462	87	AC010142	Homo sapi	AC010142	C 442	28	0.5	158207	67	AC021942	Homo sapi	
C 370	28	0.5	137557	87	AC009414	Homo sapi	AC009414	C 443	28	0.5	158276	71	AC027709	Homo sapi	
C 371	28	0.5	137604	67	AC022275	Homo sapi	AC022275	C 444	28	0.5	158297	81	AL513545	Homo sapi	
C 372	28	0.5	137658	92	HS1022J11	Human DNA	AL049765	C 445	28	0.5	158608	70	AC026834	Homo sapi	
C 373	28	0.5	137940	63	AC015612	Homo sapi	AC015612	C 446	28	0.5	158620	74	AC069522	Homo sapi	



155	29	0.6	159487	71	AC036214	Homo sapi	AC036214	Homo sapi	c 228	29	0.6	199988	68	AC023545	Homo sapi
156	29	0.6	160228	81	AL512372	Homo sapi	AL512372	Homo sapi	c 229	29	0.6	200242	61	AC010895	Homo sapi
157	29	0.6	161514	88	AC0078845	Homo sapi	AC0078845	Homo sapi	c 230	29	0.6	200594	86	AC005632	Homo sapi
158	29	0.6	161940	87	AC0078838	Homo sapi	AC0078838	Homo sapi	c 231	29	0.6	202004	97	HUAF001549	Human Chr
159	29	0.6	162339	75	AC022960	Homo sapi	AC022960	Homo sapi	c 232	29	0.6	202092	80	AL358784	Homo sapi
160	29	0.6	162470	86	AC007664	Homo sapi	AC007664	Homo sapi	c 233	29	0.6	202121	81	AL512490	Homo sapi
161	29	0.6	162922	62	AC011281	Homo sapi	AC011281	Homo sapi	c 234	29	0.6	202380	82	AL589902	Mus muscu
162	29	0.6	163282	64	AC016438	Homo sapi	AC016438	Homo sapi	c 235	29	0.6	203790	87	AC010422	Homo sapi
163	29	0.6	165872	69	AC025089	Homo sapi	AC025089	Homo sapi	c 236	29	0.6	204122	81	AL450123	Homo sapi
164	29	0.6	166297	62	AC011271	Homo sapi	AC011271	Homo sapi	c 237	29	0.6	204224	77	AC087891	Mus muscu
165	29	0.6	166884	70	AC027147	Homo sapi	AC027147	Homo sapi	c 238	29	0.6	205778	61	AC009624	Homo sapi
166	29	0.6	166884	70	AC027147	Homo sapi	AC027147	Homo sapi	c 239	29	0.6	207641	63	AC015800	Mus muscu
167	29	0.6	167460	97	UB2695	Homo sapien	UB2695	Homo sapien	c 240	29	0.6	207653	79	AL355614	Homo sapi
168	29	0.6	167629	69	AC024974	Homo sapi	AC024974	Homo sapi	c 241	29	0.6	207693	82	AL590556	Homo sapi
169	29	0.6	168043	62	AC011935	Homo sapi	AC011935	Homo sapi	c 242	29	0.6	208006	77	AC087877	Mus muscu
170	29	0.6	169677	71	AC036215	Homo sapi	AC036215	Homo sapi	c 243	29	0.6	208632	74	AC073711	Mus muscu
171	29	0.6	169819	72	AC053507	Homo sapi	AC053507	Homo sapi	c 244	29	0.6	209759	82	AP001029	Homo sapi
172	29	0.6	170128	60	AC005059	Homo sapi	AC005059	Homo sapi	c 245	29	0.6	210164	75	AC074314	Mus muscu
173	29	0.6	170197	61	AC010545	Homo sapi	AC010545	Homo sapi	c 246	29	0.6	215734	75	AC073828	Mus muscu
174	29	0.6	171049	82	AL590022	Homo sapi	AL590022	Homo sapi	c 247	29	0.6	219447	97	HUMFLNG6PD	
175	29	0.6	171140	88	AC069026	Homo sapi	AC069026	Homo sapi	c 248	29	0.6	220170	74	AC073511	Mus muscu
176	29	0.6	171490	67	AC022481	Homo sapi	AC022481	Homo sapi	c 249	29	0.6	227525	81	AL445524	Homo sapi
177	29	0.6	171549	62	AC011846	Homo sapi	AC011846	Homo sapi	c 250	29	0.6	230037	79	AL159174	Homo sapi
178	29	0.6	172241	86	AC008154	Homo sapi	AC008154	Homo sapi	c 251	29	0.6	231669	73	AC068067	Mus muscu
179	29	0.6	172332	61	AC009122	Homo sapi	AC009122	Homo sapi	c 252	29	0.6	249487	94	MMU251788	Mus muscu
180	29	0.6	173804	80	AL357835	Homo sapi	AL357835	Homo sapi	c 253	29	0.6	249487	94	MMU251835	Mus muscu
181	29	0.6	174024	66	AC021143	Homo sapi	AC021143	Homo sapi	c 254	29	0.6	262415	67	AC023143	Homo sapi
182	29	0.6	174034	63	AC013438	Homo sapi	AC013438	Homo sapi	c 255	29	0.6	265016	60	AC008383	Homo sapi
183	29	0.6	174409	61	AC010788	Homo sapi	AC010788	Homo sapi	c 256	29	0.6	269576	74	AC073684	Mus muscu
184	29	0.6	175986	67	AC023071	Homo sapi	AC023071	Homo sapi	c 257	29	0.6	271320	60	AC008758	Homo sapi
185	29	0.6	176616	60	AC007449	Homo sapi	AC007449	Homo sapi	c 258	29	0.6	324816	85	AC002368	Homo sapi
186	29	0.6	177062	61	AC009609	Homo sapi	AC009609	Homo sapi	c 259	29	0.6	327654	66	AC020816	Mus muscu
187	29	0.6	177067	64	AC016516	Homo sapi	AC016516	Homo sapi	c 260	29	0.6	340000	91	AP001710	Homo sapi
188	29	0.6	177343	65	AC018952	Homo sapi	AC018952	Homo sapi	c 261	29	0.6	340000	91	AP001728	Homo sapi
189	29	0.6	177373	83	AP002452	Homo sapi	AP002452	Homo sapi	c 262	29	0.6	340000	91	AP001754	Homo sapi
190	29	0.6	178036	82	AP001188	Homo sapi	AP001188	Homo sapi	c 263	28	0.5	284	54	G36170	STS h14a199
191	29	0.6	178109	64	AC016502	Homo sapi	AC016502	Homo sapi	c 264	28	0.5	284	54	G36170	STS h14a199
192	29	0.6	178238	64	AC016740	Homo sapi	AC016740	Homo sapi	c 265	28	0.5	511	91	BC005257	Homo sapi
193	29	0.6	178929	90	AL359252	Human DNA	AL359252	Human DNA	c 266	28	0.5	808	12	AF022741	Oryza sat
194	29	0.6	178975	67	AC023222	Homo sapi	AC023222	Homo sapi	c 267	28	0.5	880	8	AF062745	Protocler
195	29	0.6	179671	83	AP002992	Homo sapi	AP002992	Homo sapi	c 268	28	0.5	1057	9	AX044406	Sequence
196	29	0.6	179676	62	AC012221	Homo sapi	AC012221	Homo sapi	c 269	28	0.5	1155	53	CNS06TJ5	AL414615 T3 end of
197	29	0.6	179929	86	AC006270	Homo sapi	AC006270	Homo sapi	c 270	28	0.5	1215	4	AB032841	Dictyoste
198	29	0.6	180163	66	AC021510	Homo sapi	AC021510	Homo sapi	c 271	28	0.5	1330	7	SSJ000928	Scrof
199	29	0.6	180178	69	AC025287	Homo sapi	AC025287	Homo sapi	c 272	28	0.5	1432	6	HSC299386	Heteroder
200	29	0.6	180884	86	AC008018	Homo sapi	AC008018	Homo sapi	c 273	28	0.5	1436	12	AF007779	Arabirops
201	29	0.6	180942	65	AC019194	Homo sapi	AC019194	Homo sapi	c 274	28	0.5	1741	96	PMPAX6	Y09975 P.mammillat
202	29	0.6	181261	81	AL391730	Homo sapi	AL391730	Homo sapi	c 275	28	0.5	1839	89	AK022281	Homo sapi
203	29	0.6	181503	62	AC011999	Homo sapi	AC011999	Homo sapi	c 276	28	0.5	1854	91	BC005368	Homo sapi
204	29	0.6	181716	73	AC068285	Homo sapi	AC068285	Homo sapi	c 277	28	0.5	1875	97	HSU79289	Human clone
205	29	0.6	181827	67	AC022942	Homo sapi	AC022942	Homo sapi	c 278	28	0.5	3457	5	AF310886	Dictyoste
206	29	0.6	184718	87	AC018646	Homo sapi	AC018646	Homo sapi	c 279	28	0.5	3744	7	SSC251528	Sus scrof
207	29	0.6	186491	61	AC010981	Homo sapi	AC010981	Homo sapi	c 280	28	0.5	3834	14	KLKHT2GEN	247080 K.lactis KH
208	29	0.6	186812	67	AC023089	Homo sapi	AC023089	Homo sapi	c 281	28	0.5	4103	8	AF242552	Gallus ga
209	29	0.6	187182	91	CNS01DS2	BAC seque	AL121652	BAC seque	c 282	28	0.5	4163	6	DDIBP74	M29237 D.discoideu
210	29	0.6	187692	62	AC011303	Homo sapi	AC011303	Homo sapi	c 283	28	0.5	4617	96	SFAJ5398	Dictyoste
211	29	0.6	188396	82	AL590308	Homo sapi	AL590308	Homo sapi	c 284	28	0.5	18019	94	AB052890	Mus muscu
212	29	0.6	190000	78	AL139111	Homo sapi	AL139111	Homo sapi	c 285	28	0.5	35467	86	AC005570	Homo sapi
213	29	0.6	190844	75	AC073990	Homo sapi	AC073990	Homo sapi	c 286	28	0.5	39316	63	AC014425	AC005445 Drosophil
214	29	0.6	192535	70	AC027220	Homo sapi	AC027220	Homo sapi	c 287	28	0.5	40600	97	HSU101D3	285997 Human DNA s
215	29	0.6	192545	62	AC012571	Homo sapi	AC012571	Homo sapi	c 288	28	0.5	43265	77	AC087378	Homo sapi
216	29	0.6	193587	61	AC009268	Homo sapi	AC009268	Homo sapi	c 289	28	0.5	43351	85	AC004800	Homo sapi
217	29	0.6	193935	83	AP002770	Homo sapi	AP002770	Homo sapi	c 290	28	0.5	44762	85	AC002325	Homo sapi
218	29	0.6	195529	69	AC025714	Homo sapi	AC025714	Homo sapi	c 291	28	0.5	45453	12	AB022212	Arabirops
219	29	0.6	195545	81	AL392090	Homo sapi	AL392090	Homo sapi	c 292	28	0.5	50149	62	AC011955	Homo sapi
220	29	0.6	196138	79	AL353576	Homo sapi	AL353576	Homo sapi	c 293	28	0.5	51737	12	AB013391	Arabirops
221	29	0.6	196554	74	AC073128	Homo sapi	AC073128	Homo sapi	c 294	28	0.5	53000	60	AC003656	Continuation (8 of
222	29	0.6	197663	87	AC009245	Human Chr	AC009245	Human Chr	c 295	28	0.5	57196	78	AC091101	Homo sapi
223	29	0.6	197692	61	AC009554	Homo sapi	AC009554	Homo sapi	c 296	28	0.5	58238	77	AC087847	Homo sapi
224	29	0.6	198070	71	AC044909	Homo sapi	AC044909	Homo sapi	c 297	28	0.5	61373	74	AC073453	Homo sapi
225	29	0.6	198116	65	AC018440	Homo sapi	AC018440	Homo sapi	c 298	28	0.5	62850	78	AC090634	Homo sapi
226	29	0.6	198453	62	AC011359	Homo sapi	AC011359	Homo sapi	c 299	28	0.5	66260	77	AC087745	Homo sapi
227	29	0.6	198935	73	AC068573	Homo sapi	AC068573	Homo sapi	c 300	28	0.5	66594	77	AC090167	Homo sapi



9	311	5.9	54398	73	AC068224	Homo sapi	AC068224	Homo sapi	82	30	0.6	245624	80	AL359199	Homo sapi
10	297	5.6	54398	73	AC068224	Homo sapi	AC068224	Homo sapi	83	30	0.6	340000	91	AP001660	Homo sapi
11	223	4.2	72157	64	AC016449	Homo sapi	AC016449	Homo sapi	84	30	0.6	340000	91	AP001660	Homo sapi
12	163	3.1	113704	92	HS267P19	Human DNA	Z75889	Human DNA	85	29	0.6	445	13	AF329891	Pisolithu
13	163	3.1	163642	79	AL353724	Homo sapi	AL353724	Homo sapi	86	29	0.6	1511	13	AF329891	Pisolithu
14	156	3.0	163642	79	AL353724	Homo sapi	AL353724	Homo sapi	87	29	0.6	2220	6	DDCAMP1	X52888 D.discolde
15	120	2.3	69740	77	AC087748	Homo sapi	AC087748	Homo sapi	88	29	0.6	2246	94	AF017152	Mus muscu
16	100	1.9	69740	77	AC087748	Homo sapi	AC087748	Homo sapi	88	29	0.6	4548	88	AF059683	Homo sapi
17	99	1.9	150	54	G43328	Human DNA	G43328	Human DNA	89	29	0.6	6372	6	DDICNPA	M23449 Dictyosellu
18	99	1.9	150	54	G43328	Human DNA	G43328	Human DNA	90	29	0.6	6372	6	DDICNPA	M23449 Dictyosellu
19	56	1.1	340969	81	AL512630	Mus muscu	AL512630	Mus muscu	91	29	0.6	15699	85	AC003076	Human PAC
20	47	0.9	91835	92	HS26623	Human DNA	284467	Human DNA	92	29	0.6	15847	91	AP001061	Homo sapi
21	47	0.9	153665	78	AL138818	Homo sapi	AL138818	Homo sapi	93	29	0.6	43782	93	HS0J258L1	Human DNA
22	46	0.9	153665	78	AL138818	Homo sapi	AL138818	Homo sapi	94	29	0.6	44516	89	AF246630S2	Human DNA
23	41	0.8	195380	80	AL358892	Mus muscu	AL358892	Mus muscu	95	29	0.6	68631	90	AL139348	Human DNA
24	33	0.6	99030	86	AC007038	Homo sapi	AC007038	Homo sapi	96	29	0.6	71172	68	AC024313	Homo sapi
25	33	0.6	115168	77	AC084283	Homo sapi	AC084283	Homo sapi	97	29	0.6	74236	75	AC079221	Mus muscu
26	33	0.6	119395	90	AL139189	Human DNA	AL139189	Human DNA	98	29	0.6	77902	85	AC006958	Homo sapi
27	33	0.6	177551	80	AL359436	Homo sapi	AL359436	Homo sapi	99	29	0.6	83549	85	AC004752	Homo sapi
28	32	0.6	163447	90	AL354977	Human DNA	AL354977	Human DNA	100	29	0.6	85089	91	AP001432	Homo sapi
29	32	0.6	189056	68	AC024519	Homo sapi	AC024519	Homo sapi	101	29	0.6	89263	86	AC007124	Human gen
30	32	0.6	239086	69	AC024946	Homo sapi	AC024946	Homo sapi	102	29	0.6	89872	93	HSJ822J19	Human DNA
31	32	0.6	283512	80	AL365202	Homo sapi	AL365202	Homo sapi	103	29	0.6	93265	94	AP002736	Mus muscu
32	31	0.6	76986	73	AC067977	Homo sapi	AC067977	Homo sapi	104	29	0.6	95096	86	AC008384	Homo sapi
33	31	0.6	94409	67	AC022305	Homo sapi	AC022305	Homo sapi	105	29	0.6	98257	61	AC009152	Homo sapi
34	31	0.6	151088	87	AC020917	Homo sapi	AC020917	Homo sapi	106	29	0.6	99469	90	AL356961	Human DNA
35	31	0.6	156534	61	AC009700	Homo sapi	AC009700	Homo sapi	107	29	0.6	100000	91	AP000010	Homo sapi
36	31	0.6	156840	86	AC006036	Homo sapi	AC006036	Homo sapi	108	29	0.6	100000	91	AP000131	Homo sapi
37	31	0.6	159251	66	AC021659	Homo sapi	AC021659	Homo sapi	109	29	0.6	100000	91	AP000151	Homo sapi
38	31	0.6	171444	77	AC087790	Homo sapi	AC087790	Homo sapi	110	29	0.6	100000	91	AP000209	Homo sapi
39	31	0.6	171570	70	AC026454	Homo sapi	AC026454	Homo sapi	111	29	0.6	107484	92	HS4241L16	Human DNA
40	31	0.6	172105	66	AC021197	Homo sapi	AC021197	Homo sapi	112	29	0.6	108040	73	AC068150	Homo sapi
41	31	0.6	175324	74	AC090890	Canis fam	AC090890	Canis fam	113	29	0.6	110000	85	AC005079.2	Continuatio (3 of
42	31	0.6	179150	68	AC017077	Homo sapi	AC017077	Homo sapi	114	29	0.6	113916	97	HUAC003007	Human Chr
43	31	0.6	179300	66	AC020911	Homo sapi	AC020911	Homo sapi	115	29	0.6	114784	92	HS1189B24	AL030396 Human DNA
44	31	0.6	213999	88	AC051619	Homo sapi	AC051619	Homo sapi	116	29	0.6	115118	91	AP000247	Homo sapi
45	31	0.6	245324	77	AC084794	Homo sapi	AC084794	Homo sapi	117	29	0.6	120359	85	AC005389	Homo sapi
46	30	0.6	100925	83	AP002021	Homo sapi	AP002021	Homo sapi	118	29	0.6	122667	80	AL355998	Homo sapi
47	30	0.6	101539	88	AC073980	Homo sapi	AC073980	Homo sapi	119	29	0.6	126464	92	HS225E12	AL031772 Human DNA
48	30	0.6	110042	92	HS175E3	Human DNA	295113	Human DNA	120	29	0.6	127258	86	AC008562	Homo sapi
49	30	0.6	116189	92	HS98L15	Human DNA	295113	Human DNA	121	29	0.6	128915	85	AC002416	AL031281 Human Chr
50	30	0.6	128822	82	AP001260	Homo sapi	AP001260	Homo sapi	122	29	0.6	129366	92	HS224A6	AL031281 Human Chr
51	30	0.6	136769	91	AP000469	Homo sapi	AP000469	Homo sapi	123	29	0.6	130659	74	AC069368	Homo sapi
52	30	0.6	136804	91	AP000469	Homo sapi	AP000469	Homo sapi	124	29	0.6	130831	70	AC027568	Homo sapi
53	30	0.6	139391	86	AC007239	Homo sapi	AC007239	Homo sapi	125	29	0.6	134253	64	AC027568	Homo sapi
54	30	0.6	143839	85	AC004701	Homo sapi	AC004701	Homo sapi	126	29	0.6	134253	64	AC027568	Homo sapi
55	30	0.6	147519	75	AC079099	Homo sapi	AC079099	Homo sapi	127	29	0.6	134978	75	AC079004	Homo sapi
56	30	0.6	152706	79	AL355377	Homo sapi	AL355377	Homo sapi	128	29	0.6	137951	82	AL590101	Homo sapi
57	30	0.6	155450	62	AC012324	Homo sapi	AC012324	Homo sapi	129	29	0.6	138056	93	HSJ876B10	AL117352 Human DNA
58	30	0.6	155450	62	AC012324	Homo sapi	AC012324	Homo sapi	130	29	0.6	139229	78	AL139156	AL139156 Homo sapi
59	30	0.6	155722	72	AC053487	Homo sapi	AC053487	Homo sapi	131	29	0.6	139768	61	AC010931	Homo sapi
60	30	0.6	156833	86	AC006556	Homo sapi	AC006556	Homo sapi	132	29	0.6	141172	91	AL512600	AL512600 Human DNA
61	30	0.6	158037	62	AC012126	Homo sapi	AC012126	Homo sapi	133	29	0.6	143291	89	AL137792	AL137792 Human DNA
62	30	0.6	159720	88	AC068716	Homo sapi	AC068716	Homo sapi	134	29	0.6	143697	80	AL358572	Homo sapi
63	30	0.6	162584	83	AP002763	Homo sapi	AP002763	Homo sapi	135	29	0.6	146158	89	AL136136	AL136136 Human DNA
64	30	0.6	165120	89	AL138704	Human DNA	AL138704	Human DNA	136	29	0.6	146329	78	AC090889	Canis fam
65	30	0.6	166782	79	AL353896	Homo sapi	AL353896	Homo sapi	137	29	0.6	146335	63	AC012626	AC012626 Homo sapi
66	30	0.6	168305	69	AC024969	Homo sapi	AC024969	Homo sapi	138	29	0.6	147086	85	AC002472	Homo sapi
67	30	0.6	168440	79	AL162572	Homo sapi	AL162572	Homo sapi	139	29	0.6	148069	93	AC026016	AC026016 Homo sapi
68	30	0.6	168586	87	AC011497	Homo sapi	AC011497	Homo sapi	140	29	0.6	150355	99	HSJ364H10	AL078603 Human DNA
69	30	0.6	169413	80	AL358118	Homo sapi	AL358118	Homo sapi	141	29	0.6	150424	90	AL391811	Human DNA
70	30	0.6	175304	60	AC008167	Homo sapi	AC008167	Homo sapi	142	29	0.6	150697	69	AC025423	AC025423 Homo sapi
71	30	0.6	178169	90	AL391137	Human DNA	AL391137	Human DNA	143	29	0.6	150715	64	AC016258	AC016258 Homo sapi
72	30	0.6	188679	67	AC023121	Homo sapi	AC023121	Homo sapi	144	29	0.6	152452	77	AC090105	Homo sapi
73	30	0.6	189978	87	AC009075	Homo sapi	AC009075	Homo sapi	145	29	0.6	152742	70	AC027005	Homo sapi
74	30	0.6	190007	79	AL158155	Homo sapi	AL158155	Homo sapi	146	29	0.6	152794	90	AL359085	Human DNA
75	30	0.6	190793	71	AC034298	Homo sapi	AC034298	Homo sapi	147	29	0.6	153206	86	AC006356	AC006356 Homo sapi
76	30	0.6	190963	68	AC023597	Homo sapi	AC023597	Homo sapi	148	29	0.6	154835	73	AC069045	AC069045 Homo sapi
77	30	0.6	194520	90	AL356915	Human DNA	AL356915	Human DNA	149	29	0.6	155391	76	AC083801	AC083801 Homo sapi
78	30	0.6	198629	81	AL359168	Homo sapi	AL359168	Homo sapi	150	29	0.6	155414	74	AC069176	Homo sapi
79	30	0.6	199489	70	AC027505	Homo sapi	AC027505	Homo sapi	151	29	0.6	156100	62	AC011934	AC011934 Homo sapi
80	30	0.6	200000	60	AC004604	Homo sapi	AC004604	Homo sapi	152	29	0.6	156100	62	AC011934	AC011934 Homo sapi
81	30	0.6	203748	74	AC069250	Homo sapi	AC069250	Homo sapi	153	29	0.6	157485	75	AC079056	AC079056 Homo sapi
									154	29	0.6	158427	85	AC002553	Homo sapi
									154	29	0.6	159126	75	AC078795	Homo sapi



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:46:47 ; Search time 6800.24 Seconds  
(without alignments)  
11989.357 Million cell updates/sec

Title: US-09-512-581-1  
Perfect score: 5271  
Sequence: 1 ccggagagcccgagtgag.....gtacaaaaaaaaaaaaaa 5271

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 1344157 seqs, 7733874588 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_ro1.\*  
95: gb\_ro2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	80.3	5253	97	HS095825
2	4033	76.5	7444	92	HS26H231
3	3826	72.6	5110	85	AB023196
4	1126	21.4	1852	97	HS050533
5	850	16.1	137246	92	HS49J10
6	760	14.4	168487	89	AL138820
7	441	8.4	2079	89	AK026889
8	441	8.4	72157	64	AC016449



**THIS PAGE BLANK (USPTO)**







CC -1- PTM: COVALENTLY LINKED TO A SMALL RNA OF ABOUT 20 NUCLEOTIDES.  
DR PIR: A01520; TNOB01.  
DR PIR: S06426; S06426.  
KW Immunopotentialization; Thymus; Acetylation; Nuclear protein;  
KW Phosphorylation.  
FT PEPTIDE 1 28 THYMOSIN ALPHA-1.  
FT MOD\_RES 1 1 ACETYLTATION.  
FT DOMAIN 40 99 ASP/GLU-RICH (ACIDIC).  
FT MOD\_RES 7 7 PHOSPHORYLTATION (BY CK2).  
FT MOD\_RES 12 12 PHOSPHORYLTATION (BY CK2).  
FT MOD\_RES 13 13 PHOSPHORYLTATION (BY CK2).  
SQ SEQUENCE 109 AA; ECB151F0A9F99D24 CRC64;  
  
Query Match 0.5%; Score 7; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1325 EEEEEEE 1331  
DB 60 EEEEEEE 66  
  
RESULT 50  
THYA\_HUMAN  
ID THYA\_HUMAN STANDARD; PRT; 110 AA.  
AC P06454;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROTHYMOSIN ALPHA.  
GN PTMA OR TWSA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RS SEQUENCE FROM N.A.  
RX MEDLINE=87092260; PubMed=3467312;  
RA Eschenfeldt W.H., Berger S.L.;  
RT "The human prothymosin alpha gene is polymorphic and induced upon  
RT growth stimulation: evidence using a cloned cDNA";  
RT Proc. Natl. Acad. Sci. U.S.A. 83:9403-9407(1986).  
RN [2]  
RS SEQUENCE FROM N.A.  
RX MEDLINE=87067426; PubMed=3466166;  
RA Goodall G.J., Dominguez F., Horecker B.L.;  
RT "Molecular cloning of cDNA for human prothymosin alpha";  
RT Proc. Natl. Acad. Sci. U.S.A. 83:8926-8928(1986).  
RN [3]  
RS SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89214202; PubMed=2708378;  
RA Eschenfeldt W.H., Manrow R.E., Krug M.S., Berger S.L.;  
RT "Isolation and partial sequencing of the human prothymosin alpha gene  
RT family. Evidence against export of the gene products.";  
RT J. Biol. Chem. 264:7546-7555(1989).  
RN [4]  
RS SEQUENCE OF 1-14 FROM N.A.  
RX MEDLINE=93186133; PubMed=7916742;  
RA Szabo P., Panneerselvam C., Clinton M., Frangou-Lazaridis M.,  
RA Weksler D., Whittington E., Macera M.J., Grzeschik K.H.,  
RA Selvakumar A., Horecker B.L.;  
RT "Prothymosin alpha gene in humans: organization of its promoter  
RT region and localization to chromosome 2";  
RT Hum. Genet. 90:629-634(1993).  
CC -1- FUNCTION: PROTHYMOSIN ALPHA MAY MEDIATE IMMUNE FUNCTION BY  
CC CONFERRING RESISTANCE TO CERTAIN OPPORTUNISTIC INFECTIONS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: COVALENTLY LINKED TO A SMALL RNA OF ABOUT 20 NUCLEOTIDES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M67480; AAA63240.1; -  
DR EMBL; J04797; AAA63240.1; JOINED.  
DR EMBL; M67480; AAA63239.1; -  
DR EMBL; J04797; AAA63239.1; JOINED.  
DR EMBL; M14483; AAA61183.1; -  
DR EMBL; M14630; AAA61182.1; -  
DR EMBL; S56449; AAD13882.1; -  
DR PIR; A25920; TNHUA.  
DR PIR; A26654; A26654.  
DR PIR; B33356; B33356.  
DR MIM; 188390; -  
KW Immunopotentialization; Thymus; Acetylation; Nuclear protein.  
FT INIT\_MET 0 0 THYMOSIN ALPHA-1.  
FT PEPTIDE 1 28 ACETYLTATION (PROBABLE).  
FT MOD\_RES 1 1 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 41 100 MISSING (IN REF. 2).  
FT CONFLICT 39 39  
SQ SEQUENCE 110 AA; 12072 MW; 0A6609469986D272 CRC64;  
  
Query Match 0.5%; Score 7; DB 1; Length 110;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1325 EEEEEEE 1331  
DB 61 EEEEEEE 67  
  
Search completed: September 25, 2001, 10:08:03  
Job time: 154 sec



```

OX NCBI_TaxID=10116;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=87277429; PubMed=2886401;
RA Falzon M., Sanderson N., Chung S.Y.;
RT "Cloning and expression of rat homeo-box-containing sequences.";
RL Gene 54:23-32(1987).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16807; -; NOT_ANNOTATED_CDS.
CC PIR; A27471; A27471.
CC HSSP; P02833; 1HOM.
CC TRANSFAC; T01707; -.
CC InterPro; IPR001827; -.
CC IPR001356; -.
CC Pfam; PF00046; homeobox; 1.
CC PROSITE; PS00032; ANTENNAPEDIA; PARTIAL.
CC PROSITE; PS00027; HOMEBOX.1; 1.
CC PROSITE; PS00071; HOMEBOX.2; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC Transcription regulation.
CC NON_TER 1 1
CC DOMAIN 5 64 HOMEBOX.
CC DOMAIN 91 105 GLU-RICH (ACIDIC).
CC SEQUENCE 105 AA; 12552 MW; 106C1DF938F2864B CRC64;

Query Match 0.5%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1331
DB 91 EEEEEEE 97

RESULT 48
ID RL12_ARCFU STANDARD; PRT; 106 AA.
AC O28780;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L12P.
GN RPL12P OR APL1492.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervlavage A.R., Graham D.E., Kyrpides N.C.,
RA Firkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

```

---

```

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000999; AAB89748.1; -.
CC TIGR; APL1492; -.
CC InterPro; IPR001813; -.
CC Pfam; PF00428; 60s_ribosomal; 1.
CC Ribosomal protein.
CC SEQUENCE 106 AA; 11037 MW; 48956678256633B1 CRC64;

Query Match 0.5%; Score 7; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1331
DB 84 EEEEEEE 90

RESULT 49
ID THYA_BOVIN STANDARD; PRT; 109 AA.
AC P01252;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTHYMOSIN ALPHA.
GN PTMA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RE SEQUENCE.
RX MEDLINE=88339364; PubMed=2901823;
RA Panneerselvam C., Wellner D., Horecker B.L.;
RT "The amino acid sequence of bovine thymus prothymosin alpha.";
RL Arch. Biochem. Biophys. 265:454-457(1988).
RN [2]
RE SEQUENCE OF 1-28.
RX MEDLINE=79109664; PubMed=762108;
RA Low T.L.K., Goldstein A.L.;
RT "The chemistry and biology of thymosin. II. Amino acid sequence
RT analysis of thymosin alpha and polypeptide beta.";
RL J. Biol. Chem. 254:987-995(1979).
RN [3]
RE PHOSPHORYLATION BY CK2.
RX MEDLINE=93050183; PubMed=1426245;
RA Barcia M.G., Castro J.M., Jullien C.D., Gonzalez C.G., Freire M.;
RT "Prothymosin alpha is phosphorylated by casein kinase-2.";
RL FEBS Lett. 312:152-156(1992).
CC -!- FUNCTION: PROTHYMOSIN ALPHA MAY MEDIATE IMMUNE FUNCTION BY
CC CONFERRING RESISTANCE TO CERTAIN OPPORTUNISTIC INFECTIONS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.

```



RA MEDLINE-89338759; PubMed-2759245;  
RA Trompeter H.-I., Brand I.A., Soeling H.-D.;  
RA "The primary sequence of the PFK-1 inactivating zinc-binding protein  
RA as deduced from cDNA sequencing. Identity of the zinc-binding protein  
RA with rat parathymosin.";  
RL FEBS Lett. 253:63-66(1989).  
RN [2]  
RF SEQUENCE FROM N.A.  
RC STRAIN-FISCHER; TISSUE=Liver;  
RX MEDLINE-92183876; PubMed-1544455;  
RA Trompeter H.-I., Soeling H.-D.;  
RA "Cloning and characterization of a gene encoding the 11.5 kDa zinc-  
RA binding protein (parathymosin-alpha).";  
RL FEBS Lett. 298:245-248(1992).  
RN [3]  
RF SEQUENCE OF 22-101 FROM N.A.  
RC TISSUE=Spleen;  
RX MEDLINE-88239673; PubMed-3377505;  
RA Frangou-Lazaridis M., Clinton M., Goodall G.J., Horecker B.L.;  
RA "Prothymosin alpha and parathymosin: amino acid sequences deduced  
RA from the cloned rat spleen cDNAs.";  
RL Arch. Biochem. Biophys. 263:305-310(1988).  
RN [4]  
RF PRELIMINARY SEQUENCE.  
RC MEDLINE-86149281; PubMed-3456585;  
RA Komiyama T., Pan L.-X., Haritos A.A., Wideman J.W., Pan Y.-C.E.,  
RA Chang M., Rogers I., Horecker B.L.;  
RA "The primary structure of rat parathymosin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:1242-1245(1986).  
RN [5]  
RF PRELIMINARY SEQUENCE OF 1-30.  
RC MEDLINE-85140226; PubMed-3856246;  
RA Haritos A.A., Salvin S.B., Blacher R., Stein S., Horecker B.L.;  
RA "Parathymosin alpha: a peptide from rat tissues with structural  
RA homology to prothymosin alpha.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:1050-1053(1985).  
RN [6]  
RF REVISION TO 1-4.  
RC MEDLINE-88339930; PubMed-3421960;  
RA Panneerselvam C., Clinton M., Wellner D., Horecker B.L.;  
RA "Bovine parathymosin: amino acid sequence and comparison with rat  
RA parathymosin.";  
RL Biochem. Biophys. Res. Commun. 155:539-545(1988).  
CC -!- FUNCTION: PARATHYMOSIN MAY MEDIATE IMMUNE FUNCTION BY BLOCKING  
CC THE EFFECT OF PROTHYMOSIN ALPHA WHICH CONFER RESISTANCE TO  
CC CERTAIN OPPORTUNISTIC INFECTIONS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M20616; AAA42249.1; -;  
DR EMBL; M33025; AAA41810.1; -;  
DR EMBL; X64053; CAA45411.1; ALT\_SEQ.  
DR EMBL; X16481; CAA34501.1; -;  
DR PIR; B31512; B31512  
DR PIR; A05095; A05095.  
DR PIR; A22075; A22075.  
DR PIR; B32265; B32265.  
DR PIR; S05212; S05212.  
KW Immunopotentialization; Thymus; Acetylation; Zinc.  
FT INIT\_MPT 0  
FT MOD\_RES 1  
FT MOD\_RES 35  
FT DOMAIN 89  
FT DOMAIN ASP/GLU-RICH (ACIDIC).  
FT SEQUENCE 101 AA; 11428 MW; D32BE03032B85D4 CRC64;  
Query Match 0.5%; Score 7; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1325 EEEEEEE 1331  
DB 66 EEEEEEE 72  
RESULT 46  
HSP3\_MOUSE  
ID HSP3\_MOUSE STANDARD; PRT; 102 AA.  
AC Q62100;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SPERM PROTAMINE P3.  
GN PRM3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RX MEDLINE-96341725; PubMed-8720108;  
RA Schlueter G., Celik A.B., Obata R., Schllicker M., Hofferbert S.,  
RA Schlung A., Adham I.M., Engel W.;  
RA "Sequence analysis of the conserved protamine gene cluster shows that  
RA it contains a fourth expressed gene.";  
RL Mol. Reprod. Dev. 43:1-6(1996).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 247352; CAA87412.1; -;  
DR MGD; MGI:106601; Prm3.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nuclear protein.  
FT DOMAIN 46 68  
FT DOMAIN 75 78  
FT DOMAIN ASP/GLU-RICH (HIGHLY ACIDIC).  
FT POLY-LEU.  
FT SEQUENCE 102 AA; 11409 MW; E7AA59D313AF846C CRC64;  
Query Match 0.5%; Score 7; DB 1; Length 102;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1325 EEEEEEE 1331  
DB 54 EEEEEEE 60  
RESULT 47  
HXA7\_RAT  
ID HXA7\_RAT STANDARD; PRT; 105 AA.  
AC P09634;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE HOMEOBOX PROTEIN HOXA-7 (HOX-1.1) (R5) (FRAGMENT).  
GN HOXA7 OR HOXA-7.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.



CC PROXIMAL SEQUENCE ELEMENT (PSE), A NON-TATA-BOX BASAL PROMOTER  
CC ELEMENT COMMON TO THESE 2 TYPES OF GENES.  
CC -1- SUBUNIT: COMPOSED OF 5 SUBUNITS: SNAP190; SNAP50; SNAP45; SNAP43  
CC AND SNAP19.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF093593; AAC61873.1; .  
CC TRANSCRIPTION regulation. POLY-GLU.  
CC DOMAIN 86 95  
CC SEQUENCE 98 AA; 11328 MW; 4D797E35AF2D1485 CRC64;  
CC -----  
CC Query Match 0.5%; Score 7; DB 1; Length 98;  
CC Best Local Similarity 100.0%; Pred. No. 39;  
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 1325 EEEEEEE 1331  
CC |  
CC 86 EEEEEEE 92  
CC -----  
CC RESULT 43  
CC RL12\_METH STANDARD; PRT; 101 AA.  
CC AC P05394;  
CC DT 01-NOV-1988 (Rel. 09, Created)  
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)  
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
CC DE 50S RIBOSOMAL PROTEIN L12P ('A' TYPE).  
CC GN RPL12P OR MTH1682.  
CC OS Methanobacterium thermoautotrophicum.  
CC OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
CC OC Methanothermobacter.  
CC OX NCBI\_TaxID=145262;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=DELTA H;  
CC MEDLINE=98037514; PubMed=9371463;  
CC Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
CC Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
CC Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
CC Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
CC Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
CC McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
CC Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
CC "Complete genome sequence of Methanobacterium thermoautotrophicum  
CC deltaH: functional analysis and comparative genomics.";  
CC J. Bacteriol. 179:7135-7155(1997).  
CC [2]  
CC SEQUENCE OF 1-48.  
CC MEDLINE=81110590; PubMed=7006702;  
CC Matheson A.T., Yaguchi M., Balch W.E., Wolfe R.S.;  
CC "Sequence homologies in the N-terminal region of the ribosomal 'A'  
CC proteins from Methanobacterium thermoautotrophicum and Halobacterium  
CC cutirubrum.";  
CC Biochim. Biophys. Acta 626:162-169(1980).  
CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS  
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR  
CC ACCURATE TRANSLATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000926; AAB86154.1; .  
CC InterPro; IPR001813; .  
CC Pfam; PF00428; 60s\_ribosomal; 1.  
CC KW Ribosomal protein.  
CC SEQUENCE 101 AA; 10480 MW; 6C1CEA3EF121PB85 CRC64;  
CC -----  
CC Query Match 0.5%; Score 7; DB 1; Length 101;  
CC Best Local Similarity 100.0%; Pred. No. 40;  
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 1325 EEEEEEE 1331  
CC |  
CC 80 EEEEEEE 86  
CC -----  
CC RESULT 44  
CC THYP\_BOVIN STANDARD; PRT; 101 AA.  
CC ID THYP\_BOVIN  
CC AC P08814;  
CC DT 01-NOV-1988 (Rel. 09, Created)  
CC DT 01-NOV-1988 (Rel. 09, Last sequence update)  
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
CC DE PARATHYMOSIN.  
CC GN PTMS.  
CC OS Bos taurus (Bovine).  
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
CC OC Bovidae; Bovinae; Bos.  
CC OX NCBI\_TaxID=9913;  
CC [1]  
CC SEQUENCE.  
CC RX MEDLINE=88339930; PubMed=3421960;  
CC Panneerselvam C., Clinton M., Wellner D., Horecker B.L.;  
CC "Bovine parathymosin: amino acid sequence and comparison with rat  
CC parathymosin.";  
CC RL Biochem. Biophys. Res. Commun. 155:539-545(1988).  
CC -1- FUNCTION: PARATHYMOSIN MAY MEDIATE IMMUNE FUNCTION BY BLOCKING  
CC THE EFFECT OF PROTHYMOSIN ALPHA WHICH CONFER RESISTANCE TO  
CC CERTAIN OPPORTUNISTIC INFECTIONS.  
CC KW Immunopotential; Thymus; Acetylation.  
CC FT MOD\_RES 1 1 ACETYLATION.  
CC FT DOMAIN 35 89 ASP/GLU-RICH (ACIDIC).  
CC SQ SEQUENCE 101 AA; 11326 MW; 187F7D05B603000C CRC64;  
CC -----  
CC Query Match 0.5%; Score 7; DB 1; Length 101;  
CC Best Local Similarity 100.0%; Pred. No. 40;  
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 1325 EEEEEEE 1331  
CC |  
CC 66 EEEEEEE 72  
CC -----  
CC RESULT 45  
CC THYP\_RAT STANDARD; PRT; 101 AA.  
CC ID THYP\_RAT  
CC AC P04550;  
CC DT 13-AUG-1987 (Rel. 05, Created)  
CC DT 13-AUG-1987 (Rel. 05, Last sequence update)  
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
CC DE PARATHYMOSIN (ZINC-BINDING 11.5 KDA PROTEIN).  
CC GN PTMS OR ZNBP.  
CC OS Rattus norvegicus (Rat).  
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
CC OX NCBI\_TaxID=10116;  
CC [1]  
CC SEQUENCE FROM N.A.











[1] SEQUENCE FROM N.A.  
MEDLINE=97123494; PubMed=8968741;  
Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,  
Gibbons R.J.;  
"ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome."; Hum. Mol. Genet. 5:1899-1907(1996).  
[2] SEQUENCE FROM N.A.  
MEDLINE=97386582; PubMed=9244431;  
Villard L., Lossi A.M., Cardoso C., Proud V., Chiaroni P., Collea L., Schwartz C., Fontes M.;  
"Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase."; Genomics 43:149-155(1997).  
[3] SEQUENCE OF 743-2375 FROM N.A.  
MEDLINE=95179111; PubMed=7874112;  
Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V., Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E., Bianchi M.E., Consalez G.G.;  
"Cloning and characterization of a new human Xql3 gene, encoding a putative helicase."; Hum. Mol. Genet. 3:1957-1964(1994).  
[4] PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
MEDLINE=94214473; PubMed=8162050;  
Gecz J., Pollard H., Consalez G., Villard L., Stayton C., Millasseau P., Khrestchatsky M., Fontes M.;  
"Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PGK1 in Xql3.3."; Hum. Mol. Genet. 3:39-44(1994).  
[5] SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.  
MEDLINE=95211835; PubMed=7697714;  
Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
"Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome)."; Cell 80:837-845(1995).  
[6] VARIANTS ATR-X.  
MEDLINE=97467722; PubMed=9326931;  
Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B., Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K., Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F., Higgs D.R.;  
"Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain."; Nat. Genet. 17:146-148(1997).  
[7] VARIANT JM GLN-2014.  
MEDLINE=96224392; PubMed=8630485;  
Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-veber P., Munnich A., Lyonnet S.;  
"XNP mutation in a large family with Juberger-Marsidi syndrome."; Nat. Genet. 12:359-360(1996).  
-!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.  
-!- SUBCELLULAR LOCATION: NUCLEAR.  
-!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL ABNORMALITIES, AND ALPHA-THALASSEMIA.  
-!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS, MICROGENITALISM AND EARLY DEATH.  
-!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
-!- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U72938; AAB49971.1; -  
DR EMBL; U72935; AAB40699.1; -  
DR EMBL; U72904; AAB40699.1; JOINED.  
DR EMBL; U72907; AAB40699.1; JOINED.  
DR EMBL; U72908; AAB40699.1; JOINED.  
DR EMBL; U72909; AAB40699.1; JOINED.  
DR EMBL; U72910; AAB40699.1; JOINED.  
DR EMBL; U72911; AAB40699.1; JOINED.  
DR EMBL; U72912; AAB40699.1; JOINED.  
DR EMBL; U72913; AAB40699.1; JOINED.  
DR EMBL; U72914; AAB40699.1; JOINED.  
DR EMBL; U72915; AAB40699.1; JOINED.  
DR EMBL; U72916; AAB40699.1; JOINED.  
DR EMBL; U72918; AAB40699.1; JOINED.  
DR EMBL; U72919; AAB40699.1; JOINED.  
DR EMBL; U72920; AAB40699.1; JOINED.  
DR EMBL; U72921; AAB40699.1; JOINED.  
DR EMBL; U72922; AAB40699.1; JOINED.  
DR EMBL; U72923; AAB40699.1; JOINED.  
DR EMBL; U72924; AAB40699.1; JOINED.  
DR EMBL; U72925; AAB40699.1; JOINED.  
DR EMBL; U72926; AAB40699.1; JOINED.  
DR EMBL; U72927; AAB40699.1; JOINED.  
DR EMBL; U72928; AAB40699.1; JOINED.  
DR EMBL; U72929; AAB40699.1; JOINED.  
DR EMBL; U72930; AAB40699.1; JOINED.  
DR EMBL; U72931; AAB40699.1; JOINED.  
DR EMBL; U72932; AAB40699.1; JOINED.  
DR EMBL; U72933; AAB40699.1; JOINED.  
DR EMBL; U72934; AAB40699.1; JOINED.  
DR EMBL; U72935; AAB49969.1; -  
DR EMBL; U72936; AAB40700.1; JOINED.  
DR EMBL; U72937; AAB40700.1; JOINED.  
DR EMBL; U72938; AAB40700.1; JOINED.  
DR EMBL; U72939; AAB40700.1; JOINED.  
DR EMBL; U72940; AAB40700.1; JOINED.  
DR EMBL; U72941; AAB40700.1; JOINED.  
DR EMBL; U72942; AAB40700.1; JOINED.  
DR EMBL; U72943; AAB40700.1; JOINED.  
DR EMBL; U72944; AAB40700.1; JOINED.  
DR EMBL; U72945; AAB40700.1; JOINED.  
DR EMBL; U72946; AAB40700.1; JOINED.  
DR EMBL; U72947; AAB40700.1; JOINED.  
DR EMBL; U72948; AAB40700.1; JOINED.  
DR EMBL; U72949; AAB40700.1; JOINED.  
DR EMBL; U72950; AAB40700.1; JOINED.  
DR EMBL; U72951; AAB40700.1; JOINED.  
DR EMBL; U72952; AAB40700.1; JOINED.  
DR EMBL; U72953; AAB40700.1; JOINED.  
DR EMBL; U72954; AAB40700.1; JOINED.  
DR EMBL; U72955; AAB40700.1; JOINED.  
DR EMBL; U72956; AAB40700.1; JOINED.  
DR EMBL; U72957; AAB40700.1; JOINED.  
DR EMBL; U72958; AAB40700.1; JOINED.  
DR EMBL; U72959; AAB40700.1; JOINED.  
DR EMBL; U72960; AAB40700.1; JOINED.  
DR EMBL; U72961; AAB40700.1; JOINED.  
DR EMBL; U72962; AAB40700.1; JOINED.  
DR EMBL; U72963; AAB40700.1; JOINED.  
DR EMBL; U72964; AAB40700.1; JOINED.  
DR EMBL; U72965; AAB40700.1; JOINED.  
DR EMBL; U72966; AAB40700.1; JOINED.  
DR EMBL; U72967; AAB40700.1; JOINED.  
DR EMBL; U72968; AAB40700.1; JOINED.  
DR EMBL; U72969; AAB40700.1; JOINED.  
DR EMBL; U72970; AAB40700.1; JOINED.  
DR EMBL; U72971; AAB40700.1; JOINED.  
DR EMBL; U72972; AAB40700.1; JOINED.  
DR EMBL; U72973; AAB40700.1; JOINED.  
DR EMBL; U72974; AAB40700.1; JOINED.  
DR EMBL; U72975; AAB40700.1; JOINED.  
DR EMBL; U72976; AAB40700.1; JOINED.  
DR EMBL; U72977; AAB40700.1; JOINED.  
DR EMBL; U72978; AAB40700.1; JOINED.  
DR EMBL; U72979; AAB40700.1; JOINED.  
DR EMBL; U72980; AAB40700.1; JOINED.  
DR EMBL; U72981; AAB40700.1; JOINED.  
DR EMBL; U72982; AAB40700.1; JOINED.  
DR EMBL; U72983; AAB40700.1; JOINED.  
DR EMBL; U72984; AAB40700.1; JOINED.  
DR EMBL; U72985; AAB40700.1; JOINED.  
DR EMBL; U72986; AAB40700.1; JOINED.  
DR EMBL; U72987; AAB40700.1; JOINED.  
DR EMBL; U72988; AAB40700.1; JOINED.  
DR EMBL; U72989; AAB40700.1; JOINED.  
DR EMBL; U72990; AAB40700.1; JOINED.  
DR EMBL; U72991; AAB40700.1; JOINED.  
DR EMBL; U72992; AAB40700.1; JOINED.  
DR EMBL; U72993; AAB40700.1; JOINED.  
DR EMBL; U72994; AAB40700.1; JOINED.  
DR EMBL; U72995; AAB40700.1; JOINED.  
DR EMBL; U72996; AAB40700.1; JOINED.  
DR EMBL; U72997; AAB40700.1; JOINED.  
DR EMBL; U72998; AAB40700.1; JOINED.  
DR EMBL; U72999; AAB40700.1; JOINED.  
DR EMBL; U73000; AAB40700.1; JOINED.  
DR EMBL; U73001; AAB40700.1; JOINED.  
DR EMBL; U73002; AAB40700.1; JOINED.  
DR EMBL; U73003; AAB40700.1; JOINED.  
DR EMBL; U73004; AAB40700.1; JOINED.  
DR EMBL; U73005; AAB40700.1; JOINED.  
DR EMBL; U73006; AAB40700.1; JOINED.  
DR EMBL; U73007; AAB40700.1; JOINED.  
DR EMBL; U73008; AAB40700.1; JOINED.  
DR EMBL; U73009; AAB40700.1; JOINED.  
DR EMBL; U73010; AAB40700.1; JOINED.  
DR EMBL; U73011; AAB40700.1; JOINED.  
DR EMBL; U73012; AAB40700.1; JOINED.  
DR EMBL; U73013; AAB40700.1; JOINED.  
DR EMBL; U73014; AAB40700.1; JOINED.  
DR EMBL; U73015; AAB40700.1; JOINED.  
DR EMBL; U73016; AAB40700.1; JOINED.  
DR EMBL; U73017; AAB40700.1; JOINED.  
DR EMBL; U73018; AAB40700.1; JOINED.  
DR EMBL; U73019; AAB40700.1; JOINED.  
DR EMBL; U73020; AAB40700.1; JOINED.  
DR EMBL; U73021; AAB40700.1; JOINED.  
DR EMBL; U73022; AAB40700.1; JOINED.  
DR EMBL; U73023; AAB40700.1; JOINED.  
DR EMBL; U73024; AAB40700.1; JOINED.  
DR EMBL; U73025; AAB40700.1; JOINED.  
DR EMBL; U73026; AAB40700.1; JOINED.  
DR EMBL; U73027; AAB40700.1; JOINED.  
DR EMBL; U73028; AAB40700.1; JOINED.  
DR EMBL; U73029; AAB40700.1; JOINED.  
DR EMBL; U73030; AAB40700.1; JOINED.  
DR EMBL; U73031; AAB40700.1; JOINED.  
DR EMBL; U73032; AAB40700.1; JOINED.  
DR EMBL; U73033; AAB40700.1; JOINED.  
DR EMBL; U73034; AAB40700.1; JOINED.  
DR EMBL; U73035; AAB40700.1; JOINED.  
DR EMBL; U73036; AAB40700.1; JOINED.  
DR EMBL; U73037; AAB40700.1; JOINED.  
DR EMBL; U73038; AAB40700.1; JOINED.  
DR EMBL; U73039; AAB40700.1; JOINED.  
DR EMBL; U73040; AAB40700.1; JOINED.  
DR EMBL; U73041; AAB40700.1; JOINED.  
DR EMBL; U73042; AAB40700.1; JOINED.  
DR EMBL; U73043; AAB40700.1; JOINED.  
DR EMBL; U73044; AAB40700.1; JOINED.  
DR EMBL; U73045; AAB40700.1; JOINED.  
DR EMBL; U73046; AAB40700.1; JOINED.  
DR EMBL; U73047; AAB40700.1; JOINED.  
DR EMBL; U73048; AAB40700.1; JOINED.  
DR EMBL; U73049; AAB40700.1; JOINED.  
DR EMBL; U73050; AAB40700.1; JOINED.  
DR EMBL; U73051; AAB40700.1; JOINED.  
DR EMBL; U73052; AAB40700.1; JOINED.  
DR EMBL; U73053; AAB40700.1; JOINED.  
DR EMBL; U73054; AAB40700.1; JOINED.  
DR EMBL; U73055; AAB40700.1; JOINED.  
DR EMBL; U73056; AAB40700.1; JOINED.  
DR EMBL; U73057; AAB40700.1; JOINED.  
DR EMBL; U73058; AAB40700.1; JOINED.  
DR EMBL; U73059; AAB40700.1; JOINED.  
DR EMBL; U73060; AAB40700.1; JOINED.  
DR EMBL; U73061; AAB40700.1; JOINED.  
DR EMBL; U73062; AAB40700.1; JOINED.  
DR EMBL; U73063; AAB40700.1; JOINED.  
DR EMBL; U73064; AAB40700.1; JOINED.  
DR EMBL; U73065; AAB40700.1; JOINED.  
DR EMBL; U73066; AAB40700.1; JOINED.  
DR EMBL; U73067; AAB40700.1; JOINED.  
DR EMBL; U73068; AAB40700.1; JOINED.  
DR EMBL; U73069; AAB40700.1; JOINED.  
DR EMBL; U73070; AAB40700.1; JOINED.  
DR EMBL; U73071; AAB40700.1; JOINED.  
DR EMBL; U73072; AAB40700.1; JOINED.  
DR EMBL; U73073; AAB40700.1; JOINED.  
DR EMBL; U73074; AAB40700.1; JOINED.  
DR EMBL; U73075; AAB40700.1; JOINED.  
DR EMBL; U73076; AAB40700.1; JOINED.  
DR EMBL; U73077; AAB40700.1; JOINED.  
DR EMBL; U73078; AAB40700.1; JOINED.  
DR EMBL; U73079; AAB40700.1; JOINED.  
DR EMBL; U73080; AAB40700.1; JOINED.  
DR EMBL; U73081; AAB40700.1; JOINED.  
DR EMBL; U73082; AAB40700.1; JOINED.  
DR EMBL; U73083; AAB40700.1; JOINED.  
DR EMBL; U73084; AAB40700.1; JOINED.  
DR EMBL; U73085; AAB40700.1; JOINED.  
DR EMBL; U73086; AAB40700.1; JOINED.  
DR EMBL; U73087; AAB40700.1; JOINED.  
DR EMBL; U73088; AAB40700.1; JOINED.  
DR EMBL; U73089; AAB40700.1; JOINED.  
DR EMBL; U73090; AAB40700.1; JOINED.  
DR EMBL; U73091; AAB40700.1; JOINED.  
DR EMBL; U73092; AAB40700.1; JOINED.  
DR EMBL; U73093; AAB40700.1; JOINED.  
DR EMBL; U73094; AAB40700.1; JOINED.  
DR EMBL; U73095; AAB40700.1; JOINED.  
DR EMBL; U73096; AAB40700.1; JOINED.  
DR EMBL; U73097; AAB40700.1; JOINED.  
DR EMBL; U73098; AAB40700.1; JOINED.  
DR EMBL; U73099; AAB40700.1; JOINED.  
DR EMBL; U73100; AAB40700.1; JOINED.  
DR EMBL; U73101; AAB40700.1; JOINED.  
DR EMBL; U73102; AAB40700.1; JOINED.  
DR EMBL; U73103; AAB40700.1; JOINED.  
DR EMBL; U73104; AAB40700.1; JOINED.  
DR EMBL; U73105; AAB40700.1; JOINED.  
DR EMBL; U73106; AAB40700.1; JOINED.  
DR EMBL; U73107; AAB40700.1; JOINED.  
DR EMBL; U73108; AAB40700.1; JOINED.  
DR EMBL; U73109; AAB40700.1; JOINED.  
DR EMBL; U73110; AAB40700.1; JOINED.  
DR EMBL; U73111; AAB40700.1; JOINED.  
DR EMBL; U73112; AAB40700.1; JOINED.  
DR EMBL; U73113; AAB40700.1; JOINED.  
DR EMBL; U73114; AAB40700.1; JOINED.  
DR EMBL; U73115; AAB40700.1; JOINED.  
DR EMBL; U73116; AAB40700.1; JOINED.  
DR EMBL; U73117; AAB40700.1; JOINED.  
DR EMBL; U73118; AAB40700.1; JOINED.  
DR EMBL; U73119; AAB40700.1; JOINED.  
DR EMBL; U73120; AAB40700.1; JOINED.  
DR EMBL; U73121; AAB40700.1; JOINED.  
DR EMBL; U73122; AAB40700.1; JOINED.  
DR EMBL; U73123; AAB40700.1; JOINED.  
DR EMBL; U73124; AAB40700.1; JOINED.  
DR EMBL; U73125; AAB40700.1; JOINED.  
DR EMBL; U73126; AAB40700.1; JOINED.  
DR EMBL; U73127; AAB40700.1; JOINED.  
DR EMBL; U73128; AAB40700.1; JOINED.  
DR EMBL; U73129; AAB40700.1; JOINED.  
DR EMBL; U73130; AAB40700.1; JOINED.  
DR EMBL; U73131; AAB40700.1; JOINED.  
DR EMBL; U73132; AAB40700.1; JOINED.  
DR EMBL; U73133; AAB40700.1; JOINED.  
DR EMBL; U73134; AAB40700.1; JOINED.  
DR EMBL; U73135; AAB40700.1; JOINED.  
DR EMBL; U73136; AAB40700.1; JOINED.  
DR EMBL; U73137; AAB40700.1; JOINED.  
DR EMBL; U73138; AAB40700.1; JOINED.  
DR EMBL; U73139; AAB40700.1; JOINED.  
DR EMBL; U73140; AAB40700.1; JOINED.  
DR EMBL; U73141; AAB40700.1; JOINED.  
DR EMBL; U73142; AAB40700.1; JOINED.  
DR EMBL; U73143; AAB40700.1; JOINED.  
DR EMBL; U73144; AAB40700.1; JOINED.  
DR EMBL; U73145; AAB40700.1; JOINED.  
DR EMBL; U73146; AAB40700.1; JOINED.  
DR EMBL; U73147; AAB40700.1; JOINED.  
DR EMBL; U73148; AAB40700.1; JOINED.  
DR EMBL; U73149; AAB40700.1; JOINED.  
DR EMBL; U73150; AAB40700.1; JOINED.  
DR EMBL; U73151; AAB40700.1; JOINED.  
DR EMBL; U73152; AAB40700.1; JOINED.  
DR EMBL; U73153; AAB40700.1; JOINED.  
DR EMBL; U73154; AAB40700.1; JOINED.  
DR EMBL; U73155; AAB40700.1; JOINED.  
DR EMBL; U73156; AAB40700.1; JOINED.  
DR EMBL; U73157; AAB40700.1; JOINED.  
DR EMBL; U73158; AAB40700.1; JOINED.  
DR EMBL; U73159; AAB40700.1; JOINED.  
DR EMBL; U73160; AAB40700.1; JOINED.  
DR EMBL; U73161; AAB40700.1; JOINED.  
DR EMBL; U73162; AAB40700.1; JOINED.  
DR EMBL; U73163; AAB40700.1; JOINED.  
DR EMBL; U73164; AAB40700.1; JOINED.  
DR EMBL; U73165; AAB40700.1; JOINED.  
DR EMBL; U73166; AAB40700.1; JOINED.  
DR EMBL; U73167; AAB40700.1; JOINED.  
DR EMBL; U73168; AAB40700.1; JOINED.  
DR EMBL; U73169; AAB40700.1; JOINED.  
DR EMBL; U73170; AAB40700.1; JOINED.  
DR EMBL; U73171; AAB40700.1; JOINED.  
DR EMBL; U73172; AAB40700.1; JOINED.  
DR EMBL; U73173; AAB40700.1; JOINED.  
DR EMBL; U73174; AAB40700.1; JOINED.  
DR EMBL; U73175; AAB40700.1; JOINED.  
DR EMBL; U73176; AAB40700.1; JOINED.  
DR EMBL; U73177; AAB40700.1; JOINED.  
DR EMBL; U73178; AAB40700.1; JOINED.  
DR EMBL; U73179; AAB40700.1; JOINED.  
DR EMBL; U73180; AAB40700.1; JOINED.  
DR EMBL; U73181; AAB40700.1; JOINED.  
DR EMBL; U73182; AAB40700.1; JOINED.  
DR EMBL; U73183; AAB40700.1; JOINED.  
DR EMBL; U73184; AAB40700.1; JOINED.  
DR EMBL; U73185; AAB40700.1; JOINED.  
DR EMBL; U73186; AAB40700.1; JOINED.  
DR EMBL; U73187; AAB40700.1; JOINED.  
DR EMBL; U73188; AAB40700.1; JOINED.  
DR EMBL; U73189; AAB40700.1; JOINED.  
DR EMBL; U73190; AAB40700.1; JOINED.  
DR EMBL; U73191; AAB40700.1; JOINED.  
DR EMBL; U73192; AAB40700.1; JOINED.  
DR EMBL; U73193; AAB40700.1; JOINED.  
DR EMBL; U73194; AAB40700.1; JOINED.  
DR EMBL; U73195; AAB40700.1; JOINED.  
DR EMBL; U73196; AAB40700.1; JOINED.  
DR EMBL; U73197; AAB40700.1; JOINED.  
DR EMBL; U73198; AAB40700.1; JOINED.  
DR EMBL; U73199; AAB40700.1; JOINED.  
DR EMBL; U73200; AAB40700.1; JOINED.  
DR EMBL; U73201; AAB40700.1; JOINED.  
DR EMBL; U73202; AAB40700.1; JOINED.  
DR EMBL; U73203; AAB40700.1; JOINED.  
DR EMBL; U73204; AAB40700.1; JOINED.  
DR EMBL; U73205; AAB40700.1; JOINED.  
DR EMBL; U73206; AAB40700.1; JOINED.  
DR EMBL; U73207; AAB40700.1; JOINED.  
DR EMBL; U73208; AAB40700.1; JOINED.  
DR EMBL; U73209; AAB40700.1; JOINED.  
DR EMBL; U73210; AAB40700.1; JOINED.  
DR EMBL; U73211; AAB40700.1; JOINED.  
DR EMBL; U73212; AAB40700.1; JOINED.  
DR EMBL; U73213; AAB40700.1; JOINED.  
DR EMBL; U73214; AAB40700.1; JOINED.  
DR EMBL; U73215; AAB40700.1; JOINED.  
DR EMBL; U73216; AAB40700.1; JOINED.  
DR EMBL; U73217; AAB40700.1; JOINED.  
DR EMBL; U73218; AAB40700.1; JOINED.  
DR EMBL; U73219; AAB40700.1; JOINED.  
DR EMBL; U73220; AAB40700.1; JOINED.  
DR EMBL; U73221; AAB40700.1; JOINED.  
DR EMBL; U73222; AAB40700.1; JOINED.  
DR EMBL; U73223; AAB40700.1; JOINED.  
DR EMBL; U73224; AAB40700.1; JOINED.  
DR EMBL; U73225; AAB40700.1; JOINED.  
DR EMBL; U73226; AAB40700.1; JOINED.  
DR EMBL; U73227; AAB40700.1; JOINED.  
DR EMBL; U73228; AAB40700.1; JOINED.  
DR EMBL; U73229; AAB40700.1; JOINED.  
DR EMBL; U73230; AAB40700.1; JOINED.  
DR EMBL; U73231; AAB40700.1; JOINED.  
DR EMBL; U73232; AAB40700.1; JOINED.  
DR EMBL; U73233; AAB40700.1; JOINED.  
DR EMBL; U73234; AAB40700.1; JOINED.  
DR EMBL; U73235; AAB40700.1; JOINED.  
DR EMBL; U73236; AAB40700.1; JOINED.  
DR EMBL; U73237; AAB40700.1; JOINED.  
DR EMBL; U73238; AAB40700.1; JOINED.  
DR EMBL; U73239; AAB40700.1; JOINED.  
DR EMBL; U73240; AAB40700.1; JOINED.  
DR EMBL; U73241; AAB40700.1; JOINED.  
DR EMBL; U73242; AAB40700.1; JOINED.  
DR EMBL; U73243; AAB40700.1; JOINED.  
DR EMBL; U73244; AAB40700.1; JOINED.  
DR EMBL; U73245; AAB40700.1; JOINED.  
DR EMBL; U73246; AAB40700.1; JOINED.  
DR EMBL; U73247; AAB40700.1; JOINED.  
DR EMBL; U73248; AAB40700.1; JOINED.  
DR EMBL; U73249; AAB40700.1; JOINED.  
DR EMBL; U73250; AAB40700.1; JOINED.  
DR EMBL; U73251; AAB40700.1; JOINED.  
DR EMBL; U73252; AAB40700.1; JOINED.  
DR EMBL; U73253; AAB40700.1; JOINED.  
DR EMBL; U73254; AAB40700.1; JOINED.  
DR EMBL; U73255; AAB40700.1; JOINED.  
DR EMBL; U73256; AAB40700.1; JOINED.  
DR EMBL; U73257; AAB40700.1; JOINED.  
DR EMBL; U73258; AAB40700.1; JOINED.  
DR EMBL; U73259; AAB40700.1; JOINED.  
DR EMBL; U73260; AAB40700.1; JOINED.  
DR EMBL; U73261; AAB40700.1; JOINED.  
DR EMBL; U73262; AAB40700.1; JOINED.  
DR EMBL; U73263; AAB40700.1; JOINED.  
DR EMBL; U73264; AAB40700.1; JOINED.  
DR EMBL; U73265; AAB40700.1; JOINED.  
DR EMBL; U73266; AAB40700.1; JOINED.  
DR EMBL; U73267; AAB40700.1; JOINED.  
DR EMBL; U73268; AAB40700.1; JOINED.  
DR EMBL; U73269; AAB40700.1; JOINED.  
DR EMBL; U73270; AAB40700.1; JOINED.  
DR EMBL; U73271; AAB40700.1; JOINED.  
DR EMBL; U73272; AAB40700.1; JOINED.  
DR EMBL; U73273; AAB40700.1; JOINED.  
DR EMBL; U73274; AAB40700.1; JOINED.  
DR EMBL; U73275; AAB40700.1; JOINED.  
DR EMBL; U73276; AAB40700.1; JOINED.  
DR EMBL; U73277; AAB40700.1; JOINED.  
DR EMBL; U73278; AAB40700.1; JOINED.  
DR EMBL; U73279; AAB40700.1; JOINED.  
DR EMBL; U73280; AAB40700.1; JOINED.  
DR EMBL; U73281; AAB40700.1; JOINED.  
DR EMBL; U73282; AAB40700.1; JOINED.  
DR EMBL; U73283; AAB40700.1; JOINED.  
DR EMBL; U73284; AAB40700.1; JOINED.  
DR EMBL; U73285; AAB40700.1; JOINED.  
DR EMBL; U73286; AAB40700.1; JOINED.  
DR EMBL; U73287; AAB40700.1; JOINED.  
DR EMBL; U73288; AAB40700.1; JOINED.  
DR EMBL; U73289; AAB40700.1; JOINED.  
DR EMBL; U73290; AAB40700.1; JOINED.  
DR EMBL; U73291; AAB40700.1; JOINED.  
DR EMBL; U73292; AAB40700.1; JOINED.  
DR EMBL; U73293; AAB40700.1; JOINED.  
DR EMBL; U73294; AAB40700.1; JOINED.  
DR EMBL; U73295; AAB40700.1; JOINED.  
DR EMBL; U73296; AAB40700.1; JOINED.  
DR EMBL; U73297; AAB40700.1; JOINED.  
DR EMBL; U73298; AAB40700.1; JOINED.  
DR EMBL; U73299; AAB40700.1; JOINED.  
DR EMBL; U73300; AAB40700.1; JOINED.  
DR EMBL; U73301; AAB40700.1; JOINED.  
DR EMBL; U73302; AAB40700.1; JOINED.  
DR EMBL; U73303; AAB40700.1; JOINED.  
DR EMBL; U73304; AAB40700.1; JOINED.  
DR EMBL; U73305; AAB407



```

EMBL; M32698; AAA41659.1; -.
DR PIR; S06006; S06006.
DR PIR; A02989; A02989.
DR HSP; P08799; 1MND.
DR InterPro; IPR000048; -.
DR InterPro; IPR001609; -.
DR InterPro; IPR002928; -.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 839
FT DOMAIN 840 1935
FT DOMAIN 780 802
FT DOMAIN 840 1935
FT NP_BIND 178 185
FT DOMAIN 655 677
FT DOMAIN 757 771
FT MOD_RES 129 129
FT MOD_RES 695 695
FT MOD_RES 705 705
FT CONFLICT 1529 1531
FT CONFLICT 1731 1731
FT CONFLICT 1784 1784
FT CONFLICT 1851 1851
FT CONFLICT 1858 1858
FT CONFLICT 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;
SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1935;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKQLE 564
DB 1527 EKIRKQLE 1534

RESULT 36
ID CVAL_DROME STANDARD; PRT; 2248 AA.
AC P32870;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CA2+/CALMODULIN-RESPONSIVE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (RUTABAGA PROTEIN).
DE RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=Head;
RA MEDLINE=92154664; PubMed=1739965;
RA Levin L.R., Han P.-L., Hwang P.M., Feinstein P.G., Davis R.L.,
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
RT Ca2+/calmodulin-responsive adenylyl cyclase."
RT Cell 68:479-489(1992).
CC -!- FUNCTION: THIS IS A MEMBRANE-BOUND, CALMODULIN-SENSITIVE ADENYLYL
CC CYCLASE. INACTIVATION OF THIS CYCLASE LEADS TO A LEARNING AND
CC MEMORY DEFECT.
CC -!- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -!- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN AND G PROTEIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.

EMBL; M32698; AAA41659.1; -.
DR PIR; S06006; S06006.
DR PIR; A02989; A02989.
DR HSP; P08799; 1MND.
DR InterPro; IPR000048; -.
DR InterPro; IPR001609; -.
DR InterPro; IPR002928; -.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 839
FT DOMAIN 840 1935
FT DOMAIN 780 802
FT DOMAIN 840 1935
FT NP_BIND 178 185
FT DOMAIN 655 677
FT DOMAIN 757 771
FT MOD_RES 129 129
FT MOD_RES 695 695
FT MOD_RES 705 705
FT CONFLICT 1529 1531
FT CONFLICT 1731 1731
FT CONFLICT 1784 1784
FT CONFLICT 1851 1851
FT CONFLICT 1858 1858
FT CONFLICT 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;
SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1935;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKQLE 564
DB 1527 EKIRKQLE 1534

RESULT 36
ID CVAL_DROME STANDARD; PRT; 2248 AA.
AC P32870;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CA2+/CALMODULIN-RESPONSIVE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (RUTABAGA PROTEIN).
DE RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=Head;
RA MEDLINE=92154664; PubMed=1739965;
RA Levin L.R., Han P.-L., Hwang P.M., Feinstein P.G., Davis R.L.,
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
RT Ca2+/calmodulin-responsive adenylyl cyclase."
RT Cell 68:479-489(1992).
CC -!- FUNCTION: THIS IS A MEMBRANE-BOUND, CALMODULIN-SENSITIVE ADENYLYL
CC CYCLASE. INACTIVATION OF THIS CYCLASE LEADS TO A LEARNING AND
CC MEMORY DEFECT.
CC -!- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -!- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN AND G PROTEIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.

EMBL; M81887; AAA28844.1; -.
DR PIR; D42088; D42088.
DR HSP; P19754; 1AWK.
DR FlyBase; FBgn0003301; rut.
DR InterPro; IPR001054; -.
DR Pfam; PF00211; guanylate_cyc; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
DR PROSITE; PS01125; GUANYLATE_CYCLASES_2; 2.
KW Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
FT DOMAIN 1 41
FT TRANSMEM 42 60
FT TRANSMEM 65 84
FT TRANSMEM 101 115
FT TRANSMEM 122 142
FT TRANSMEM 152 174
FT TRANSMEM 186 206
FT TRANSMEM 207 705
FT DOMAIN 706 726
FT TRANSMEM 730 750
FT TRANSMEM 770 791
FT TRANSMEM 792 813
FT TRANSMEM 814 834
FT TRANSMEM 842 867
FT TRANSMEM 889 2248
FT DOMAIN 318 334
FT DOMAIN 1013 1029
FT DOMAIN 515 530
FT DOMAIN 569 602
FT DOMAIN 1278 1297
FT DOMAIN 1767 1810
FT DOMAIN 2025 2040
FT DOMAIN 2200 2241
FT CARBOHYD 800 800
FT CARBOHYD 807 807
FT MUTAGEN 1026 1026
SQ SEQUENCE 2248 AA; 248899 MW; E459C718BE018868 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 2248;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LLLSVLPQ 276
DB 238 LLLSVLPQ 245

RESULT 37
ID ATRX_HUMAN STANDARD; PRT; 2375 AA.
AC P46100; P51068; Q15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
DE ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```



AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
-!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
-!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).  
-!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L12104; AAA62313.1; -  
EMBL; X07273; CAA30256.1; -  
PIR; A28298; A28298.  
HSSP; P08799; LMND.  
InterPro; IPR000048; -  
InterPro; IPR001609; -  
InterPro; IPR002928; -  
Pfam; PF00612; IQ; 1.  
Pfam; PF01576; Myosin\_tail; 1.  
Pfam; PF00063; myosin\_head; 1.  
PRINTS; PR00193; MYOSINHEAVY.  
PROSITE; PS00096; IQ; 1.  
Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Alkylation; Multigene family.  
DOMAIN 1 838 GLOBULAR HEAD (S1).  
DOMAIN 839 1934 RODLIKE TAIL (S2 AND LMM DOMAINS).  
DOMAIN 779 801 IQ.  
DOMAIN 839 1934 COILED COIL (POTENTIAL).  
NP\_BIND 177 184 ATP.  
DOMAIN 654 676 ACTIN-BINDING.  
DOMAIN 756 770 ACTIN-BINDING.  
MOD\_RES 128 128 METHYLATION (TRI-) (POTENTIAL).  
MOD\_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).  
MOD\_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).  
CONFLICT 966 966 D -> E (IN REF. 2).  
CONFLICT 978 978 T -> TE (IN REF. 2).  
CONFLICT 986 986 E -> Q (IN REF. 2).  
CONFLICT 1008 1014 DLQAEED -> ALQARKT (IN REF. 2).  
CONFLICT 1057 1057 D -> Y (IN REF. 2).  
CONFLICT 1060 1060 L -> V (IN REF. 2).  
CONFLICT 1095 1095 D -> N (IN REF. 2).  
CONFLICT 1217 1217 E -> D (IN REF. 2).  
CONFLICT 1271 1271 D -> N (IN REF. 2).  
CONFLICT 1327 1327 T -> A (IN REF. 2).  
CONFLICT 1358 1358 C -> R (IN REF. 2).  
CONFLICT 1504 1504 L -> V (IN REF. 2).  
CONFLICT 1537 1537 M -> L (IN REF. 2).  
CONFLICT 1556 1556 N -> K (IN REF. 2).  
SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1934;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKQLE 564  
|||||||  
DB 1526 EKIRKQLE 1533

RESULT 35  
ID MYSB\_RAT  
AC P02564; STANDARD; PRT; 1935 AA.

DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.  
GN MYH7.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart.  
RX MEDLINE=90016823; PubMed=2798112;  
RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
RT "Complete nucleotide sequence of full length cDNA for rat beta  
cardiac myosin heavy chain.";  
RL Nucleic Acids Res. 17:7529-7530(1989).  
RN [2]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=90133919; PubMed=2614840;  
RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.  
Comparisons suggest a molecular basis for functional differences.";  
RL J. Mol. Biol. 210:665-671(1989).  
RN [3]  
RP SEQUENCE OF 1524-1935 FROM N.A.  
RX MEDLINE=82220036; PubMed=7045682;  
RA Mahdavi V., Periasamy M., Nadal-Ginard B.;  
RT "Molecular characterization of two myosin heavy chain genes expressed  
in the adult heart.";  
RL Nature 297:659-664(1982).  
RN [4]  
RP SEQUENCE OF 1871-1935 FROM N.A.  
RC STRAIN=Wistar; TISSUE=Heart;  
RX MEDLINE=85179510; PubMed=6241892;  
RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;  
RT "Cardiac myosin heavy chain isozymic transitions during development  
and under pathological conditions are regulated at the level of mRNA  
availability.";  
RL Eur. Heart J. 5:181-191(1984).  
CC -!- FUNCTION: MUSCLE CONTRACTION.  
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE  
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
SUBFRAGMENT (S2).  
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE  
MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.  
CC -!- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN  
THE RAT, TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY  
CHAINS.  
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,  
WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE  
CONSERVED.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X15939; CAA34065.1; -  
EMBL; J00752; AAA41654.1; -



between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; AB018284; BAA34461.1; -.
CC DR EMBL; AF078035; AAD16006.1; -.
CC DR EMBL; AJ006412; CAA07018.1; -.
CC DR InterPro: IPR000795; -.
CC DR Pfam; PF00009; GTP_EFTU; 1.
CC KW Initiation factor; Protein biosynthesis; GTP-binding.
CC FT NP_BIND 638 645 GTP (BY SIMILARITY).
CC FT DOMAIN 39 50 POLY-LYS.
CC FT DOMAIN 94 99 POLY-LYS.
CC FT DOMAIN 138 142 POLY-ASP.
CC FT DOMAIN 313 322 POLY-LYS.
CC FT DOMAIN 353 356 POLY-LYS.
CC FT DOMAIN 361 364 POLY-GLU.
CC FT DOMAIN 491 496 POLY-GLU.
CC FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).
CC FT MUTAGEN 640 640 V->G: LOSS OF ACTIVITY IN VIVO. RETAINS
CC FT MUTAGEN 706 706 FULL ACTIVITY IN VITRO.
CC FT MUTAGEN 706 706 H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND
CC FT MUTAGEN 706 706 IN VITRO.
CC FT MUTAGEN 759 759 H->Q: LOSS OF ACTIVITY IN VIVO. PARTIAL
CC FT MUTAGEN 759 759 ACTIVITY IN VITRO.
CC FT MUTAGEN 759 759 D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND
CC FT CONFLICT 64 64 G -> E (IN REF. 2).
CC FT CONFLICT 180 180 I -> M (IN REF. 2).
CC FT CONFLICT 256 256 K -> R (IN REF. 2).
CC FT CONFLICT 522 522 K -> T (IN REF. 2).
CC FT CONFLICT 549 549 E -> V (IN REF. 2).
CC FT CONFLICT 549 549 E -> V (IN REF. 2).
CC FT CONFLICT 669 669 G -> W (IN REF. 2).
CC FT SEQUENCE 1220 AA; 138754 MW; 61F0435138B262E7 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1220;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 531 PEEEEEE 538
|||||||

RESULT 33
ID GBFL_HUMAN STANDARD; PRT; 1859 AA.
AC Q92538;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GOLGI-SPECIFIC BREFELDIN A-RESISTANCE GUANINE NUCLEOTIDE EXCHANGE
DE FACTOR 1 (BFA-RESISTANT GEF 1).
GN GBFL OR KIA0248.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047539; PubMed=9828135;
RA Mansour S.J., Herbrick J.-A., Scherer S.W., Melancon P.;
RT "Human GBFL is a ubiquitously expressed gene of the sec7 domain family
RT mapping to 10q24.";
RL Genomics 54:323-327(1998).
RN [2]
RP SEQUENCE OF 169-1859 FROM N.A.
RX TISSUE=Bone marrow;
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC MEDLINE=97191544; PubMed=9039502;

```

```

RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF5. PROMOTES
CC THE ACTIVATION OF ARF5 THROUGH REPLACEMENT OF GDP WITH GTP (BY
CC SIMILARITY).
CC -!- ENZYME REGULATION: INHIBITED BY BREFELDIN A (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AF068755; AAD15903.1; -.
CC DR EMBL; D87435; BAA13379.1; -.
CC DR MIM; 603698; -.
CC DR HSSP; Q99418; IPR000904; -.
CC DR InterPro: IPR000904; -.
CC DR Pfam; PF01369; Sec7; 1.
CC KW Guanine-nucleotide releasing factor.
CC FT DOMAIN 696 884 SEC7.
CC FT SEQUENCE 1859 AA; 206444 MW; 5015D2BD70009CFA CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1859;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 DDEDEGVP 641
DB 1477 DDEDEGVP 1484
|||||||

RESULT 34
ID MYSB_MESAU STANDARD; PRT; 1934 AA.
AC P13540; Q60540;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLB; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
RT heavy chain gene from Syrian hamster.";
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
RT heavy chain.";
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

```



-!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX. COMPONENT OF BOTH  
 BASEMENT MEMBRANES AND OTHER CONNECTIVE TISSUES.  
 -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA AND OVARY.  
 -!- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.  
 -!- SIMILARITY: CONTAINS 11 EGF-LIKE DOMAINS.

---

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; X82494; CAA57876.1; -.  
HSP; P07204; IFGD.  
MIM; 135821; -.  
InterPro; IPR000020; -.  
InterPro; IPR000152; -.  
InterPro; IPR000561; -.  
InterPro; IPR001881; -.  
Pfam; PF01821; ANATO; 2.  
Pfam; PF00008; EGF; 7.  
PROSITE; PS00010; ASX\_HYDROXYL; 5.  
PROSITE; PS01177; ANAPHYLATOXIN.1; 3.  
PROSITE; PS01178; ANAPHYLATOXIN.2; 3.  
PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
PROSITE; PS01186; EGF\_2; 5.  
PROSITE; PS01187; EGF\_CA; 9.  
Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain;  
Calcium-binding; Repeat.

SIGNAL	1	27	POTENTIAL.
CHAIN	28	1184	FIBULIN-2.
DOMAIN	28	444	N.
DOMAIN	28	177	SUBDOMAIN NA (CYS-RICH).
DOMAIN	178	444	SUBDOMAIN NB (CYS-FREE).
DOMAIN	445	480	ANAPHYLATOXIN-LIKE 1.
DOMAIN	488	519	ANAPHYLATOXIN-LIKE 2.
DOMAIN	521	553	ANAPHYLATOXIN-LIKE 3.
DOMAIN	604	645	EGF-LIKE 1, CALCIUM-BINDING.
DOMAIN	679	718	EGF-LIKE 2, CALCIUM-BINDING.
DOMAIN	719	763	EGF-LIKE 3, CALCIUM-BINDING.
DOMAIN	764	809	EGF-LIKE 4, CALCIUM-BINDING.
DOMAIN	810	857	EGF-LIKE 5, CALCIUM-BINDING.
DOMAIN	858	900	EGF-LIKE 6, CALCIUM-BINDING.
DOMAIN	901	942	EGF-LIKE 7, CALCIUM-BINDING.
DOMAIN	943	981	EGF-LIKE 8, CALCIUM-BINDING.
DOMAIN	982	1024	EGF-LIKE 9, CALCIUM-BINDING.
DOMAIN	1025	1069	EGF-LIKE 10, CALCIUM-BINDING.
DOMAIN	1070	1184	DOMAIN III.
DISULFID	445	472	BY SIMILARITY.
DISULFID	446	479	BY SIMILARITY.
DISULFID	459	480	BY SIMILARITY.
DISULFID	489	518	BY SIMILARITY.
DISULFID	502	519	BY SIMILARITY.
DISULFID	521	545	BY SIMILARITY.
DISULFID	522	552	BY SIMILARITY.
DISULFID	535	553	BY SIMILARITY.
DISULFID	608	620	BY SIMILARITY.
DISULFID	616	629	BY SIMILARITY.
DISULFID	631	644	BY SIMILARITY.
DISULFID	683	693	BY SIMILARITY.
DISULFID	689	702	BY SIMILARITY.
DISULFID	704	717	BY SIMILARITY.
DISULFID	723	736	BY SIMILARITY.
DISULFID	730	745	BY SIMILARITY.
DISULFID	751	762	BY SIMILARITY.
DISULFID	768	781	BY SIMILARITY.
DISULFID	775	790	BY SIMILARITY.
DISULFID	796	808	BY SIMILARITY.
DISULFID	814	827	BY SIMILARITY.
DISULFID	821	836	BY SIMILARITY.
DISULFID	843	856	BY SIMILARITY.

```

FT DISULFID 862 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 899 BY SIMILARITY.
FT DISULFID 905 917 BY SIMILARITY.
FT DISULFID 913 926 BY SIMILARITY.
FT DISULFID 928 941 BY SIMILARITY.
FT DISULFID 947 956 BY SIMILARITY.
FT DISULFID 952 965 BY SIMILARITY.
FT DISULFID 967 980 BY SIMILARITY.
FT DISULFID 986 998 BY SIMILARITY.
FT DISULFID 994 1007 BY SIMILARITY.
FT DISULFID 1009 1023 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1036 1051 BY SIMILARITY.
FT DISULFID 1056 1068 BY SIMILARITY.
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1035 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 1184 AA; 126543 MW; C48490A5F9EC5D CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1184;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 EEEEEER 1332
    |||||
Db 280 EEEEEER 287

RESULT 32
IF2P_HUMAN
ID IC IP2P_HUMAN STANDARD; PRT; 1220 AA.
AC O60841; O95805;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DE TRANSLATION INITIATION FACTOR IF-2.
GN KIAA0741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.;"
RL DNA Res. 5:277-286(1998).
RN [2]
RS SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=99218282; PubMed=10200264;
RA Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;
RT "Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2.;"
RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
RN [3]
RS SEQUENCE OF 833-1220 FROM N.A..
RA Schiichi F., Bertocco E., Lanfranchi G., Valle G.;
RT "Finding homologues between human and yeast.;"
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO FUNCTION ALONG WITH EIF-2 (IF-2 FAMILY).
CC -1- SIMILARITY: BELONGS TO EIF-2 (IF-2 FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration







(CA(2+)-INHIBITABLE ADENYLYL CYCLASE).  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=92357702; PubMed=1379717;  
Yoshimura M., Cooper D.M.F.;  
"Cloning and expression of a Ca(2+)-inhibitable adenylyl cyclase from  
NCB-20 cells";  
Proc. Natl. Acad. Sci. U.S.A. 89:6716-6720(1992).  
[2]  
SEQUENCE OF 10-1165 FROM N.A.  
MEDLINE=93076707; PubMed=1332848;  
Premont R.T., Jacobowitz O., Ivengar R.;  
"Lowered responsiveness of the catalyst of adenylyl cyclase to  
stimulation by GS in heterologous desensitization: a role for  
adenosine 3',5'-monophosphate-dependent phosphorylation";  
Endocrinology 131:2774-2784(1992).  
-!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL  
CYCLASE.  
-!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
-!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR  
CONCENTRATION RANGE.  
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
-!- TISSUE SPECIFICITY: MOST ABUNDANT IN HEART BUT WEAKLY DETECTABLE  
IN BRAIN, INTESTINE, LUNG, AND SPLEEN.  
-!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.  
-!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE  
FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M93422; AAA37174.1; -;  
EMBL; M96653; AAA37182.1; -;  
PIR; A46187; A46187.  
HSP; P19754; IAWK.  
MGD; MGI:87917; Adcy6.  
InterPro; IPR001054; -;  
Pfam; PF00211; guanylate\_cyc; 2.  
PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 2.  
PROSITE; PS00125; GUANYLATE\_CYCLASES\_2; 2.  
Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.  
DOMAIN 1 149  
CYTOPLASMIC (POTENTIAL).  
TRANSMEM 150 166  
TRANSMEM 179 195  
TRANSMEM 212 228  
TRANSMEM 237 253  
TRANSMEM 257 273  
TRANSMEM 287 303  
DOMAIN 304 670  
CYTOPLASMIC (POTENTIAL).  
TRANSMEM 671 688  
TRANSMEM 699 715  
TRANSMEM 740 756  
DOMAIN 757 816  
TRANSMEM 817 833  
TRANSMEM 836 852  
TRANSMEM 894 910  
TRANSMEM 911 1165  
CYTOPLASMIC (POTENTIAL).  
CARBOHYD 277 277  
CARBOHYD 790 790  
CARBOHYD 875 875  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
K -> N (IN REF. 2).  
G -> A (IN REF. 2).  
CONFLICT 76 76  
CONFLICT 47 47  
GR -> RAG (IN REF. 2).  
CONFLICT 508 509

FT CONFLICT 737 737 V -> G (IN REF. 2).  
FT CONFLICT 881 881 L -> Q (IN REF. 2).  
FT CONFLICT 990 990 V -> M (IN REF. 2).  
SQ SEQUENCE 1165 AA; 130318 MW; 24EE1BB45DF1E87E CRC64;  
  
Query Match 0.6%; Score 8; DB 1; Length 1165;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 269 LLSVLPPQ 276  
Db 339 LLSVLPPQ 346  
  
RESULT 29  
CYA6\_RAT STANDARD; PRT; 1166 AA.  
ID CYA6\_RAT  
AC Q03343;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ADENYLYL CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)  
DE (CA(2+)-INHIBITABLE ADENYLYL CYCLASE).  
GN ADCY6.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93077589; PubMed=1332969;  
RA Krupinski J., Lehman T.C., Frankenfield C.D., Zwaagstra J.C.,  
Watson P.A.;  
RT "Molecular diversity in the adenylyl cyclase family. Evidence for  
eight forms of the enzyme and cloning of type VI";  
RL J. Biol. Chem. 267:24858-24862(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=93028552; PubMed=1409703;  
RA Premont R.T., Chen J., Ma H.-W., Ponnappalli M., Ivengar R.;  
RT "Two members of a widely expressed subfamily of hormone-stimulated  
adenylyl cyclases";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:9809-9813(1992).  
CC -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL  
CYCLASE.  
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
CC -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR  
CONCENTRATION RANGE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.  
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE  
FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L01115; AAA40676.1; -;  
EMBL; M96160; AAA40678.1; ALT\_INIT.  
PIR; A45145; A45145.  
HSP; P19754; IAWK.  
InterPro; IPR001054; -;  
Pfam; PF00211; guanylate\_cyc; 2.  
PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 2.  
PROSITE; PS00125; GUANYLATE\_CYCLASES\_2; 2.  
Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.  
DOMAIN 1 149  
CYTOPLASMIC (POTENTIAL).



15-DEC-1998 (Rel. 37, Last annotation update)  
 SUCROSE-PHOSPHATE SYNTHASE 2 (EC 2.4.1.14) (UDP-GLUCOSE-FRUCTOSE-  
 PHOSPHATE GLUCOSYLTRANSFERASE 2).  
 SPS2.  
 Craterostigma plantaginenum.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 Lamiales; Scrophulariaceae; Craterostigma.  
 NCBI\_TaxID=4153;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=974511773; PubMed=9306694;  
 Ingram J., Chandler J.W., Gallagher L., Salamini F., Bartels D.;  
 "Analysis of cDNA clones encoding sucrose-phosphate synthase in  
 relation to sugar interconversions associated with dehydration in the  
 resurrection plant Craterostigma plantaginenum Hochst.;"  
 Plant Physiol. 113:113-121(1997).  
 -!- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN  
 THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND  
 THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF  
 PHOTASSIMILATES OUT OF THE LEAF.  
 -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE 6-PHOSPHATE = UDP +  
 SUCROSE 6-PHOSPHATE.  
 -!- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND  
 MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.  
 -!- PATHWAY: SUCROSE SYNTHESIS PATHWAY.  
 -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER (BY SIMILARITY).  
 -!- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR  
 ENZYME FUNCTION (BY SIMILARITY).  
 -!- SIMILARITY: SOME, WITH SUCROSE SYNTHASES.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 EMBL: Y11795; CAA72491.1; -;  
 InterPro: IPR001296; -;  
 Pfam: PF00534; Glycosyltransferase; Phosphorylation; Multigene family.  
 Transferrase; Glycosyltransferase; Phosphorylation; Multigene family.  
 DOMAIN 245 248 POLY-SER.  
 DOMAIN 256 264 POLY-GLU.  
 DOMAIN 787 790 POLY-ARG.  
 SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;  
 -----  
 Query Match 0.6%; Score 8; DB 1; Length 1081;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1324 PEEEEEE 1331  
 |||||  
 DB 257 PEEEEEE 264  
 -----  
 RESULT 27  
 CYA6\_CANFA STANDARD; PRT; 1165 AA.  
 ID CYA6\_CANFA  
 AC P30804;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)  
 DE (CA(2+))-INHIBITABLE ADENYLYL CYCLASE).  
 DE ADCY6.  
 GN Canis familiaris (Dog).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 [1]

SEQUENCE FROM N.A.  
 TISSUE=Heart;  
 MEDLINE=92409599; PubMed=1528892;  
 Katsushika S., Chen L., Kawabe J., Nilakantan R., Halnon N.J.,  
 Homcy C.J., Ishikawa Y.;  
 "Cloning and characterization of a sixth adenylyl cyclase isoform:  
 types V and VI constitute a subgroup within the mammalian adenylyl  
 cyclase family.;"  
 Proc. Natl. Acad. Sci. U.S.A. 89:8774-8778(1992).  
 -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL  
 CYCLASE.  
 -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
 -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR  
 CONCENTRATION RANGE.  
 -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.  
 -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE  
 FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 EMBL: M94968; -; NOT\_ANNOTATED\_CDS.  
 PIR: A46180; A46180.  
 HSP: P19754; IAWK.  
 InterPro: IPR001054; -;  
 Pfam: PF00211; guanylate\_cyc; 2.  
 PROSITE: PS00452; GUANYLATE\_CYCLASES\_1; 2.  
 PROSITE: PS0125; GUANYLATE\_CYCLASES\_2; 2.  
 KW Lyase; cAMP synthesis; Transmembrane; Glycoprotein; Duplication.  
 FT DOMAIN 1 149  
 FT TRANSMEM 150 166  
 FT TRANSMEM 179 195  
 FT TRANSMEM 212 228  
 FT TRANSMEM 237 253  
 FT TRANSMEM 257 273  
 FT TRANSMEM 287 303  
 FT DOMAIN 304 670  
 FT TRANSMEM 671 688  
 FT TRANSMEM 699 715  
 FT TRANSMEM 740 756  
 FT DOMAIN 757 816  
 FT TRANSMEM 817 833  
 FT TRANSMEM 836 852  
 FT TRANSMEM 894 910  
 FT TRANSMEM 911 1165  
 FT CARBOHYD 790 790  
 FT CARBOHYD 875 875  
 FT SEQUENCE 1165 AA; 130323 MW; BA9D2D329120615E CRC64;  
 SQ  
 -----  
 Query Match 0.6%; Score 8; DB 1; Length 1165;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 269 LLLSVLPQ 276  
 |||||  
 DB 339 LLLSVLPQ 346  
 -----  
 RESULT 28  
 CYA6\_MOUSE STANDARD; PRT; 1165 AA.  
 ID CYA6\_MOUSE  
 AC Q01341;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)



RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,  
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehman H., Reinhardt R., Yaspo M.-L.,  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
CC -!- FUNCTION: MAY DEUBIQUITINATE ONE OR MORE CRITICAL PROTEINS THAT  
CC ARE INVOLVED IN THE CONDENSATION OF MITOTIC CHROMOSOMES, POSSIBLY  
CC ACTING SELECTIVELY ON HISTONES H2A AND H2B, THE MAJOR  
CC UBIQUITINATED PROTEINS OF CHROMATIN. IT IS ABLE TO DEUBIQUITINATE  
CC HISTONE H2A IN VITRO, THE PHOSPHORYLATED FORM OF THE PROTEIN IS  
CC ALSO ENZYMATICALLY ACTIVE.  
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =  
CC UBIQUITIN + A THIOL.  
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL THE TISSUES EXAMINED INCLUDING  
CC FETAL BRAIN, LUNG, LIVER, KIDNEY, AND ADULT HEART, BRAIN,  
CC PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.  
CC -!- PTM: PHOSPHORYLATED AT THE ONSET OF MITOSIS AND  
CC DEPHOSPHORYLATED DURING THE METAPHASE/ANAPHASE TRANSITION.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS  
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF126736; AAD20949.1; -;  
DR EMBL; AL163249; CAB90432.1; -;  
DR MEROPS; C19.021; -;  
DR MIM; 604735; -;  
DR InterPro; IPR001394; -;  
DR Pfam; PF00442; UCH-1; 1.  
DR Pfam; PF00443; UCH-2; 1.  
DR PROSITE; PS00972; UCH\_2\_1; 1.  
DR PROSITE; PS00973; UCH\_2\_2; 1.  
DR PROSITE; PS0235; UCH\_2\_3; 1.  
KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family;  
FT Phosphorylation.  
FT ACT\_SITE 205 205 BY SIMILARITY.  
FT ACT\_SITE 750 750 BY SIMILARITY.  
FT ACT\_SITE 758 758 BY SIMILARITY.  
FT MUTAGEN 205 205 C->S: UNABLE TO DEUBIQUITINATE H2A/H2B.  
SQ SEQUENCE 823 AA; 93570 MW; C7D4175649BA3E31 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 823;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 DQDSSEK 53  
DQ 412 DQDSSEK 419

RESULT 25  
IF2P\_YEAST STANDARD; PRT; 1002 AA.  
AC P39730;  
DT 01-FEB-1995 (Rel. 31, Created)  
DI 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TRANSLATION INITIATION FACTOR IF-2.

GN YAL035W OR FUN12.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=95249563; PubMed=7731988;  
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
RA Storms R.K.;  
RT "The nucleotide sequence of chromosome I from Saccharomyces  
RT cerevisiae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94357438; PubMed=8076820;  
RA Sutcliffe P., Shafer B.K., Strathern J.N., Hughes S.H.;  
RT "Isolation, identification and characterization of the FUN12 gene of  
RT Saccharomyces cerevisiae.";  
RL Gene 146:209-213(1994).  
RN [3]  
RP REVISIONS  
RC STRAIN=S288C / AB972;  
RA Vo D.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP CHARACTERIZATION  
RX MEDLINE=98289357; PubMed=9624054;  
RA Choi S.K., Lee J.H., Zoll W.L., Merrick W.C., Dever T.E.;  
RT "Promotion of met-tRNA<sup>Met</sup> binding to ribosomes by yif2, a bacterial  
RT IF2 homolog in yeast.";  
RL Science 280:1757-1760(1998).  
CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING  
CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO  
CC FUNCTION ALONG WITH EIF-2.  
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U12980; AAC04996.1; -;  
DR EMBL; L29389; AAA57228.1; ALT\_SEQ.  
DR SGD; S0000033; FUN12.  
DR InterPro; IPR000795; -;  
DR Pfam; PF00009; GTP\_EFTU; 1.  
KW Initiation factor; Protein biosynthesis; GTP-binding.  
FT DOMAIN 361 371 POLY-GLU.  
FT NP\_BIND 412 419 GTP (BY SIMILARITY).  
SQ SEQUENCE 1002 AA; 112268 MW; 1A496195DAE1C283 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1002;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
DQ 365 EEEEEER 372

RESULT 26  
SPS2\_CRAPL STANDARD; PRT; 1081 AA.  
AC 004933;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)



-!- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS FARNESYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE ENVELOPE INTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.

-!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.

-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

THIS IS A B TYPE LAMIN.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X16705; CAA34677.1; -  
EMBL; M35153; AAC96023.1; ALT\_SEQ.  
PIR; S07720; S07720.  
SWISS-2DPAGE; P14733; MOUSE.  
MGD; MGI:96795; Lmmb1.  
InterPro; IPR001322; -  
InterPro; IPR001664; -  
Pfam; PF00932; IF\_tail; 1.  
Pfam; PF00038; filament; 1.  
PROSITE; PS00226; IF; 1.  
Intermediate filament; Heptad repeat pattern; Coiled coil;  
Nuclear protein; Lipoprotein; Prenylation; Phosphorylation.  
INIT\_MET 0  
DOMAIN 1 34 HEAD.  
DOMAIN 35 386 ROD.  
DOMAIN 387 587 TAIL.  
DOMAIN 35 69 COIL 1A.  
DOMAIN 70 81 LINKER 1.  
DOMAIN 82 215 COIL 1B.  
DOMAIN 216 243 LINKER 2.  
DOMAIN 244 386 COIL 2.  
DOMAIN 415 420 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
DOMAIN 552 560 GLU-RICH (HIGHLY ACIDIC, COULD BE INVOLVED IN CHROMATIN BINDING).  
LIPID 584 584 FARNESYL (BY SIMILARITY).  
SEQUENCE 587 AA; 66753 MW; 222E87CDF053E4B7 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 587;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
DB 551 PEEEEEE 558  
|||||

RESULT 23  
APPL\_MOUSE STANDARD; PRT; 653 AA.  
AC Q03157;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE ANKYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).  
GN APLP1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=93066322; PubMed=1279693;

RA Wasco W., Bupp K., Magendantz M., Gusella J.F., Tanzi R.E., Solomon F.;  
RT "Identification of a mouse brain cDNA that encodes a protein related to the Alzheimer disease-associated amyloid beta protein precursor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP IS SECRETED.  
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L04538; AAA37247.1; -  
DR PIR; A46362; A46362.  
DR MGD; MGI:88046; Aplp1.  
DR InterPro; IPR001868; -  
DR PRINTS; PR00203; AMYLOIDA4.  
DR PROSITE; PS00319; A4\_EXTRA; 1.  
DR PROSITE; PS00320; A4\_INTRA; 1.  
KW Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 653 AMYLOID-LIKE PROTEIN 1.  
FT DOMAIN 21 583 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 584 606 POTENTIAL.  
FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 643 646 CLATHRIN-BINDING (POTENTIAL).  
FT DOMAIN 263 271 POLY-GLU.  
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 653;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332  
DB 265 EEEEEEE 272  
|||||

RESULT 24  
UBPG\_HUMAN STANDARD; PRT; 823 AA.  
AC Q9Y3T5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16)  
DE (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M).  
GN USP16.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99178974; PubMed=10077596;  
RA Cai S.-Y., Babbitt R.W., Marchesi V.T.;  
RT "A mutant deubiquitinating enzyme (Ubp-M) associates with mitotic chromosomes and blocks cell division";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2828-2833(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20289799; PubMed=10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,



Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1330  
DQ 462 APEEEEE 469

RESULT 21  
LAMI\_HUMAN  
ID LAMI\_HUMAN STANDARD; PRT; 585 AA.  
AC P20700;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE LAMIN B1.  
GN LMNB1 OR LMNB2 OR LMNB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=90220602; PubMed=2325650;  
RA Pollard K.M., Chan E.K.L., Grant B.J., Sullivan K.F., Tan E.M.,  
RA Glass C.A.;  
RT "In vitro posttranslational modification of lamin B cloned from a  
RT human T-cell line.";  
RL Mol. Cell. Biol. 10:2164-2175(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96044426; PubMed=7557986;  
RA Lin F., Worman H.J.;  
RT "Structural organization of the human gene (LMNB1) encoding nuclear  
RT lamin B1.";  
RL Genomics 27:230-236(1995).  
RN [3]  
RP ISOPRENOID.  
RA MEDLINE=90062174; PubMed=2684976;  
RA Farnsworth C.C., Wolda S.L., Gelb M.H., Glomset J.A.;  
RT "Human lamin B contains a farnesylated cysteine residue.";  
RL J. Biol. Chem. 264:20422-20429(1989).  
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
CC AND MAY ALSO INTERACT WITH CHROMATIN.  
CC -!- SUBCELLULAR LOCATION: NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR  
CC MEMBRANE.  
CC -!- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS  
CC FARNESYLATION AND PHOSPHORYLATION, INCREASED PHOSPHORYLATION OF  
CC THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY  
CC PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.  
CC -!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY  
CC CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND  
CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,  
CC RESPECTIVELY.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC THIS IS A B TYPE LAMIN.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M34458; AAC36162.1; -  
EMBL; L37747; AAC37575.1; -  
EMBL; L37737; AAC37575.1; JOINED.  
EMBL; L37738; AAC37575.1; JOINED.  
EMBL; L37739; AAC37575.1; JOINED.  
EMBL; L37740; AAC37575.1; JOINED.

DR EMBL; L37741; AAC37575.1; JOINED.  
DR EMBL; L37742; AAC37575.1; JOINED.  
DR EMBL; L37743; AAC37575.1; JOINED.  
DR EMBL; L37744; AAC37575.1; JOINED.  
DR EMBL; L37745; AAC37575.1; JOINED.  
DR EMBL; L37746; AAC37575.1; JOINED.  
DR PIR; A34707; A34707.  
DR SWISS-2DPAGE; P20700; HUMAN.  
DR Aarhus/Ghent-2DPAGE; 7510; IEF.  
DR MIM; 150340; -  
DR InterPro; IPR001322; -  
DR InterPro; IPR001664; -  
DR Pfam; PF00932; IF\_tail; 1.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Heptad repeat pattern; Coiled coil;  
KW Nuclear protein; Lipoprotein; Prenylation; Phosphorylation.  
FT INIT\_MET 0  
FT DOMAIN 1 33 HEAD.  
FT DOMAIN 34 385 ROD.  
FT DOMAIN 386 585 TAIL.  
FT DOMAIN 34 68 COIL 1A.  
FT DOMAIN 69 80 LINKER 1.  
FT DOMAIN 81 214 COIL 1B.  
FT DOMAIN 215 242 LINKER 2.  
FT DOMAIN 243 385 COIL 2.  
FT DOMAIN 414 419 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 551 558 GLU-RICH (HIGHLY ACIDIC, COULD BE  
FT INVOLVED IN CHROMATIN BINDING).  
FT LIPID 582 582 FARNESYL.  
SQ SEQUENCE 585 AA; 66277 MW; 0802EDAC33CBF998 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 585;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
DQ 550 PEEEEEE 557

RESULT 22  
LAMI\_MOUSE  
ID LAMI\_MOUSE STANDARD; PRT; 587 AA.  
AC P14733;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE LAMIN B1.  
GN LMNB1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=89210899; PubMed=3243285;  
RA Hoeger T.H., Krohne G., Franke W.W.;  
RT "Amino acid sequence and molecular characterization of murine lamin B  
RT as deduced from cDNA clones.";  
RL Eur. J. Cell Biol. 47:283-290(1988).  
RN [2]  
RP REVISIONS.  
RA Hoeger T.H.;  
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
CC AND MAY ALSO INTERACT WITH CHROMATIN.  
CC -!- SUBCELLULAR LOCATION: NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR  
CC MEMBRANE.



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U70158; AAC50135.1; -

MIM: 601603; -

HSSP: P15034; LBKM.

InterPro: IPR000980; -

Pfam: PF00017; SH2; 1.

PROSITE: PS50001; SH2; 1.

SH2 domain; Phosphorylation.

POLY-GLU.

POLY-PRO.

DOMAIN 133 136

DOMAIN 198 201

DOMAIN 422 530

SEQUENCE 533 AA; 60188 MW; C5D22F31D36200C8 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 533;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 SSPLPGKK 1197

|||||

297 SSPLPGKK 304

RESULT 19

INVO\_AOTTR

ID INVO\_AOTTR STANDARD; PRT; 544 AA.

AC P24708;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE INVOLUCRIN.

GN IVL.

OS Aotus trivirgatus (Night monkey) (Douroucoulis).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.

OX NCBI\_TaxID=9505;

RN [1]

RE SEQUENCE FROM N.A.

RA MEDLINE=90014142; PubMed=2507864;

RS Tseng H., Green H.;

RT "The involucrin gene of the owl monkey: origin of the early region.";

RI Mol. Biol. Evol. 6:460-468(1989).

CC -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS

CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO

CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE

CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.

CC -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND

CC OTHER STRATIFIED SQUAMOUS EPITHELIA.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M25313; AAA5375.1; -

HSSP: P80220; 1DIP.

InterPro: IPR000354; -

Pfam: PF00904; Involutrin; 42.

PROSITE: PS00795; INVOLUCRIN; 1.

Keratinocyte; Repeat.

SEQUENCE 544 AA; 63927 MW; 2A02ABA5E1499F9D CRC64;

Query Match 0.6%; Score 8; DB 1; Length 544;

Query Match

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 SQELLDTV 199

|||||

Db 16 SQELLDTV 23

RESULT 20

ID NFL\_XENLA

AC P35616; STANDARD; PRT; 544 AA.

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NEUROFILAMENT TRIPLET L PROTEIN (NEUROFILAMENT LIGHT POLYPEPTIDE) (NF-

DE L).

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=92356194; PubMed=1494944;

RA Charnas L.R., Szaro B.G., Gainer H.;

RT "Identification and developmental expression of a novel low molecular

RT weight neuronal intermediate filament protein expressed in Xenopus

RT laevis.";

RL J. Neurosci. 12:3010-3024(1992).

CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,

CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.

CC -!- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE

CC NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL

CC INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM

CC FILAMENTS

CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M86654; AAA83018.1; -

DR PIR: B44841; B44841.

DR InterPro: IPR001684; -

DR Pfam: PF00038; filament; 1.

DR PROSITE: PS00226; IF; 1.

KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone.

FT DOMAIN 1 87

FT HEAD.

FT ROD.

FT DOMAIN 88 390

FT TAIL.

FT DOMAIN 391 544

FT COIL 1A.

FT DOMAIN 88 119

FT LINKER 1.

FT DOMAIN 120 132

FT COIL 1B.

FT DOMAIN 133 228

FT LINKER 12.

FT DOMAIN 229 246

FT COIL 2A.

FT DOMAIN 247 265

FT LINKER 2.

FT DOMAIN 266 274

FT COIL 2B.

FT DOMAIN 275 390

FT TAIL, SUBDOMAIN A.

FT DOMAIN 391 435

FT TAIL, SUBDOMAIN B (ACIDIC).

FT DOMAIN 436 544

FT GLU-RICH.

FT DOMAIN 441 538

FT POLY-GLU.

FT DOMAIN 464 469

FT SEQUENCE 544 AA; 61861 MW; 76D911B896E97201 CRC64;

QY



-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X06832; CAA29988.1; --  
 PIR; A23996; A23996.  
 PIR; S00291; S00291.  
 PIR; S02543; S02543.  
 InterPro; IPR001819; --  
 InterPro; IPR001990; --  
 Pfam; PF01271; Granin; 1.  
 PRINTS; PR00659; CHROMOGRANIN.  
 PROSITE; PS00422; GRANINS\_1; 1.  
 PROSITE; PS00423; GRANINS\_2; 1.  
 Signal; Amidation; Glycoprotein; Calcium-binding.  
 SIGNAL 18  
 CHAIN 19 466  
 CHROMOGRANIN A.  
 PEPTIDE 19 146  
 BETA-GRANIN.  
 PEPTIDE 281 332  
 PANGREASTATIN (POTENTIAL).  
 PEPTIDE 361 374  
 WE-14.  
 DOMAIN 92 110  
 POLY-GLN.  
 DOMAIN 231 249  
 POLY-GLU.  
 DOMAIN 344 350  
 POLY-GLU.  
 MOD\_RES 332 332  
 AMIDATION (PROBABLE).  
 DISULFID 35 56  
 BY SIMILARITY.  
 SEQUENCE 466 AA; 52024 MW; 05D135FFA657C48C CRC64;

Query Match 0.6%; Score 8; DB 1; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 |||||  
 DR 344 EEEEEER 351

## RESULT 17

NIFK\_RHISN  
 ID NIFK\_RHISN STANDARD; PRT; 513 AA.  
 AC P13067;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NITROGENASE MOLYBDENUM-IRON PROTEIN BETA CHAIN (EC 1.18.6.1)  
 DE (NITROGENASE COMPONENT I) (DINITROGENASE).  
 GN (NIFK1 OR Y4VM) AND (NIFK2 OR Y4XC).  
 OS Rhizobium sp. (strain NGR234).  
 OC Plasmid sym pNGR234a.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
 RL Nature 387:394-401(1997).  
 RN [2]  
 RP SEQUENCE OF 132-195 FROM N.A.  
 RC STRAIN=ANU 240;  
 RX MEDLINE=89306671; PubMed=2744485;  
 RA Badenoch-Jones J., Holton T.A., Morrison C.M., Scott K.F., Shine J.;  
 RT "Structural and functional analysis of nitrogenase genes from the  
 RT broad-host-range Rhizobium strain ANU240.";  
 RL Gene 77:141-153(1989).  
 CC -!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE

CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE  
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.  
 CC -!- CATALYTIC ACTIVITY: 8 REDUCED FERREDOXIN + 8 H(+) + N(2) + 16 ATP  
 CC = 8 OXIDIZED FERREDOXIN + 2 NH(3) + 16 ADP + 16 ORTHOPHOSPHATE.  
 CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS  
 CC 30-32 FE, 2 MO, AND INORGANIC SULFUR.  
 CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL; M26963; AAA26327.1; --  
 EMBL; AE000102; AAB91901.1; --  
 EMBL; AE000105; AAB91925.1; --  
 PIR; PS0046; PS0046.  
 HSSP; P07329; IN2C.  
 InterPro; IPR000318; --  
 InterPro; IPR000510; --  
 Pfam; PF00148; Oxidored\_nitro; 1.  
 PROSITE; PS00699; NITROGENASE\_1\_1; 1.  
 PROSITE; PS00090; NITROGENASE\_1\_2; 1.  
 KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur; Plasmid;  
 KW Multigene family.  
 SQ SEQUENCE 513 AA; 57302 MW; 41631040335541AE CRC64;

Query Match 0.6%; Score 8; DB 1; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 LRLLTTL 858  
 |||||  
 DB 481 LRLLTTL 488

## RESULT 18

LCP2\_HUMAN  
 ID LCP2\_HUMAN STANDARD; PRT; 533 AA.  
 AC Q13094;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE LYMPHOCYTE CYTOSOLIC PROTEIN 2 (SH2 DOMAIN-CONTAINING LEUCOCYTE  
 DE PROTEIN OF 76 KDA) (SLP-76 TYROSINE PHOSPHOPROTEIN) (SLP76).  
 GN LCP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Leukemia;  
 RX MEDLINE=95221345; PubMed=7706237;  
 RA Jackman J.K., Motto D.G., Sun Q., Tanemoto M., Turck C.W., Peltz G.A.,  
 RA Koretzky G.A., Findell P.R.;  
 RT "Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein  
 RT associated with Grb2 in T cells.";  
 RL J. Biol. Chem. 270:7029-7032(1995).  
 CC -!- FUNCTION: INVOLVED IN T CELL ANTIGEN RECEPTOR MEDIATED SIGNALING.  
 CC -!- SUBUNIT: INTERACTS WITH THE ADAPTER PROTEINS GRB2 AND Fyb.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, THYMUS, AND  
 CC PERIPHERAL BLOOD LEUKOCYTES. HIGHLY EXPRESSED ALSO IN T CELL AND  
 CC MONOCYTIC CELL LINES. EXPRESSED AT LOWER LEVEL IN B CELL LINES.  
 CC NOT DETECTED IN FIBROBLAST OR NEUROBLASTOMA CELL LINES.  
 CC -!- PTM: PHOSPHORYLATED AFTER T-CELL RECEPTOR ACTIVATION.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -----



EMBL; X07547; CAA30421.1; -  
DR EMBL; X06707; CAA29892.1; -  
DR EMBL; J03321; AAA91569.1; -  
DR EMBL; M19487; AAB02586.1; ALT\_FRAME.  
DR EMBL; Y00505; CAA58558.1; ALT\_FRAME.  
DR PIR; S01921; S01921.  
DR PIR; S01022; S01022.  
DR PIR; S01023; S01023.  
DR PIR; S06324; S06324.  
DR InterPro; IPR001198; -  
DR Pfam; PF00772; DnaB; 1.  
KW DNA replication; DNA-binding; ATP-binding; Helicase; Primosome;  
Plasmid.  
FT NP\_BIND 225 232 ATP (POTENTIAL).  
FT VARIAT 55 55 Y -> H (IN PLASMID PCHL1).  
FT VARIAT 188 188 A -> V (IN PLASMID PCHL1).  
FT VARIAT 189 189 T -> A (IN PLASMID PCHL1 AND PCTT1).  
FT VARIAT 200 200 A -> E (IN PLASMID PCHL1).  
FT VARIAT 200 200 A -> G (IN PLASMID PCTT1).  
FT VARIAT 267 267 I -> V (IN PLASMID PCTT1).  
SQ SEQUENCE 451 AA; 51456 MW; 2CD73A152C72A1F1 CRC64;  
Query Match 0.6%; Score 8; DB 1; Length 451;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1169 RIKGRDLS 1176  
DB 145 RIKGRDLS 152  
RESULT 15  
CMGA\_MOUSE STANDARD; PRT; 463 AA.  
AC P26339;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CHROMOGRANIN A PRECURSOR (CGA) [CONTAINS: PANCREASTATIN; BETA-GRANIN;  
DE WE-14].  
GN CHGA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91302337; PubMed=2071596;  
RA Wu H.J., Rozansky D.J., Farmer R.J., Gill B.M., O'Connor D.T.;  
RT "Structure and function of the chromogranin A gene. Clues to  
RT evolution and tissue-specific expression."  
RL J. Biol. Chem. 266:13130-13134(1991).  
CC -!- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN  
CC RELEASE FROM THE PANCREAS.  
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY  
CC GRANULES.  
CC -!- PTM: CGA IS O-GLYCOSYLATED.  
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
CC -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN  
CC FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; M64278; AAA37457.1; -  
DR PIR; A39868; A39868.  
DR MGD; MGI:88394; Chga.

DR InterPro; IPR001819; -  
DR InterPro; IPR001990; -  
DR Pfam; PF01271; Granin; 1.  
DR PRINTS; PR00659; CHROMOGRANIN.  
DR PROSITE; PS00422; GRANINS\_1; 1.  
DR PROSITE; PS00423; GRANINS\_2; 1.  
KW Signal; Amidation; Glycoprotein; Calcium-binding.  
FT SIGNAL 1 18  
FT CHAIN 19 463 CHROMOGRANIN A.  
FT PEPTIDE 19 151 BETA-GRANIN (BY SIMILARITY).  
FT PEPTIDE 276 329 PANCREASTATIN (BY SIMILARITY).  
FT PEPTIDE 358 371 WE-14 (BY SIMILARITY).  
FT DOMAIN 103 116 POLY-GLN.  
FT DOMAIN 236 245 POLY-GLU.  
FT DOMAIN 340 347 POLY-GLU.  
FT MOD\_RES 329 329 AMIDATION (PROBABLE).  
FT DISULFID 35 56 BY SIMILARITY.  
SQ SEQUENCE 463 AA; 51789 MW; 1AB3C5FF433C39E4 CRC64;  
Query Match 0.6%; Score 8; DB 1; Length 463;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1325 EEEEEER 1332  
DB 341 EEEEEER 348  
RESULT 16  
CMGA\_RAT STANDARD; PRT; 466 AA.  
AC P10354;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE CHROMOGRANIN A PRECURSOR (CGA) [CONTAINS: PANCREASTATIN; BETA-GRANIN;  
DE WE-14].  
GN CHGA.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88112232; PubMed=2828116;  
RA Iacangelo A., Okayama H., Eiden L.E.;  
RT "Primary structure of rat chromogranin A and distribution of its  
RT mRNA."  
RL FEBS Lett. 227:115-121(1988).  
RN [2]  
RP SEQUENCE OF 29-466 FROM N.A.  
RX TISSUE=Pancreas;  
RX MEDLINE=88312980; PubMed=3044825;  
RA Huton J.C., Nielsen E., Kasten W.;  
RT "The molecular cloning of the chromogranin A-like precursor of beta-  
RT granin and pancreastatin from the endocrine pancreas."  
RL FEBS Lett. 236:269-274(1988).  
RN [3]  
RP SEQUENCE OF 19-32.  
RX MEDLINE=85285598; PubMed=3896848;  
RA Huton J.C., Hansen F., Peshavaria M.;  
RT "Beta-granins: 21 kDa co-secreted peptides of the insulin granule  
RT closely related to adrenal medullary chromogranin A."  
RL FEBS Lett. 188:336-340(1985).  
CC -!- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN  
CC RELEASE FROM THE PANCREAS.  
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY  
CC GRANULES.  
CC -!- PTM: CGA IS O-GLYCOSYLATED.  
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
CC -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN  
CC FAMILY.



-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AE000999; AAB89749.1; -

TIGR: AF1491; -

InterPro: IPR001790; -

Pfam: PF00466; Ribosomal\_L110; 1.

Ribosomal protein.

SEQUENCE 339 AA; 37096 MW; F8BEBACBBB710D0 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1324 PEEEEEE 1331

|||||||

317 PEEEEEE 324

#### RESULT 13

ID YSE2.CAEEL STANDARD; PRT; 418 AA.

AC Q09936;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 47.4 KDA PROTEIN C53C9.2 IN CHROMOSOME X.

GN C53C9.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Bentley D.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RI -!- SIMILARITY: BELONGS TO THE CALPONIN FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U28734; AAB52603.1; -

WormPep; C53C9.2; C201863.

InterPro; IPR000357; -

Pfam; PF00402; calponin; 4.

PROSITE; PS01052; CALPONIN; 1.

Hypothetical protein; Repeat.

REPEAT 50 75

REPEAT 94 119

REPEAT 182 210

REPEAT 276 301

REPEAT 366 418

DOMAIN GLU-RICH.

DOMAIN 366 379

DOMAIN 388 418

POLY-GLU.

SEQUENCE 418 AA; 47446 MW; DA195883CD6E6D52 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 418;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||||

Db 369 PEEEEEE 376

#### RESULT 14

GPID\_CHLTR

ID GPID\_CHLTR STANDARD; PRT; 451 AA.

AC P10555; P08781; P22445;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE PROBABLE PLASMID REPLICATIVE DNA HELICASE (DNAB-LIKE PROTEIN)

DE (VRULENCE PLASMID PROTEIN PGPI-D) (PROTEIN P-3).

OS Chlamydia trachomatis.

OC Plasmid pLGV440, plasmid pCHL1, and plasmid pCTT1.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=813;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=L2/434/BU; PLASMID=PLGV440;

RX MEDLINE=89013895; PubMed=2845228;

RA Comanducci M., Ricci S., Ratti G.;

RT "The structure of a plasmid of Chlamydia trachomatis believed to be

RT required for growth within mammalian cells.";

RL Mol. Microbiol. 2:531-538(1988).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=LL/440/LN; PLASMID=PLGV440;

RX MEDLINE=88233998; PubMed=2836808;

RA Hatt C., Ward M.B., Clarke I.N.;

RT "Analysis of the entire nucleotide sequence of the cryptic plasmid of

RT Chlamydia trachomatis serovar L1. Evidence for involvement in DNA

RT replication.";

RL Nucleic Acids Res. 16:4053-4067(1988).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=GO/86 / SEROTYPE D; PLASMID=pCHL1;

RX MEDLINE=90301796; PubMed=2194229;

RA Comanducci M., Ricci S., Cevenini R., Ratti G.;

RT "Diversity of the Chlamydia trachomatis common plasmid in biovars

RT with different pathogenicity.";

RL Plasmid 23:149-154(1990).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=SEROTYPE B; PLASMID=pCTT1;

RX MEDLINE=88171106; PubMed=3444859;

RA Sriprakash K.S., Macavoy E.S.;

RT "Characterization and sequence of a plasmid from the trachoma biovar

RT of Chlamydia trachomatis.";

RL Plasmid 18:205-214(1987).

[5]

RP SEQUENCE OF 1-407 FROM N.A.

RC STRAIN=SEROTYPE B; PLASMID=pCTT1;

RX MEDLINE=88096599; PubMed=2827127;

RA Sriprakash K.S., Macavoy E.S.;

RT "A gene for dnaB like protein in chlamydial plasmid.";

RL Nucleic Acids Res. 15:10596-10596(1987).

CC -!- FUNCTION: REQUIRED FOR GROWTH WITHIN MAMMALIAN CELLS.

CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PLASMID PLGV440.

CC -!- SIMILARITY: BELONGS TO THE DNAB SUBFAMILY OF HELICASES.

CC -!- CAUTION: REF.4 AND 5 SEQUENCES DIFFER FROM THAT SHOWN DUE TO A

CC FRAMESHIFT IN POSITION 389.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----



RA Wilber J.F.;  
RT "Cloning and structure of human genomic DNA and hypothalamic cDNA  
RL encoding human prepro thyrotropin-releasing hormone.";  
RI Mol. Endocrinol. 4:551-556(1990).  
CC -!- FUNCTION: TRH FUNCTIONS AS A REGULATOR OF THE BIOSYNTHESIS OF TSH  
CC IN THE ANTERIOR PITUITARY GLAND AND AS A NEUROTRANSMITTER/  
CC NEUROMODULATOR IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.  
CC -!- TISSUE SPECIFICITY: HYPOTHALAMUS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M63582; AAA36480.1; -;  
CC EMBL; M63581; AAA36480.1; JOINED.  
CC PIR; A34550; A34550.  
CC MIM; 275120; -;  
CC Amidation; Hormone; Repeat; Hypothalamus; Signal;  
CC Cleavage on pair of basic residues.  
CC SIGNAL 1 724 POTENTIAL.  
CC CHAIN 25 242 PROTHYROLIBERIN.  
CC PEPTIDE 84 86 THYROLIBERIN.  
CC PEPTIDE 114 116 THYROLIBERIN.  
CC PEPTIDE 135 137 THYROLIBERIN.  
CC PEPTIDE 152 154 THYROLIBERIN.  
CC PEPTIDE 186 188 THYROLIBERIN.  
CC PEPTIDE 227 229 THYROLIBERIN.  
CC MOD\_RES 84 84 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 86 86 AMIDATION (G-87 PROVIDE AMIDE GROUP).  
CC MOD\_RES 114 114 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 116 116 AMIDATION (G-117 PROVIDE AMIDE GROUP).  
CC MOD\_RES 135 135 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 137 137 AMIDATION (G-138 PROVIDE AMIDE GROUP).  
CC MOD\_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 154 154 AMIDATION (G-155 PROVIDE AMIDE GROUP).  
CC MOD\_RES 186 186 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 188 188 AMIDATION (G-189 PROVIDE AMIDE GROUP).  
CC MOD\_RES 227 227 PYRROLIDONE CARBOXYLIC ACID.  
CC MOD\_RES 229 229 AMIDATION (G-230 PROVIDE AMIDE GROUP).  
CC SEQUENCE 242 AA; 27404 MW; 8C0F9D915B32F29F CRC64;  
CC -----  
CC Query Match 0.6%; Score 8; DB 1; Length 242;  
CC Best Local Similarity 100.0%; Pred. No. 9.2;  
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC 1325 EEEEEER 1332  
CC |  
CC 169 EEEEEER 176  
CC -----  
CC RESULT 11  
CC YNFB\_YEAST STANDARD; PRT; 316 AA.  
CC AC P53947;  
CC DT 01-OCT-1996 (Rel. 34, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
CC DE HYPOTHETICAL 35.0 KDA PROTEIN IN ARP5-OMP2 INTERGENIC REGION.  
CC YNLO58C OR N2433 OR YNLO2433C.  
CC OS Saccharomyces cerevisiae (Baker's yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
CC NCBI\_TaxID=4932;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=S288C / FY1676;  
CC MEDLINE=96021608; PubMed=8533472;  
CC Bergez P., Doignon F., Crouzet M.;

RT "The sequence of a 44 420 bp fragment located on the left arm of  
RT chromosome XIV from Saccharomyces cerevisiae.";  
RL Yeast 11:967-974(1995).  
RN -----  
RP ERRATUM.  
RX MEDLINE=97060022; PubMed=8904343;  
RA Bergez P., Doignon F., Crouzet M.;  
RL Yeast 12:297-297(1996).  
CC -!- SIMILARITY: TO YEAST YIL117C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U12141; AAA99653.1; -;  
CC EMBL; 271334; CAA95932.1; -;  
CC SGD; S0005003; YNL058C.  
CC KW Hypothetical protein; Transmembrane.  
CC TRANSMEM 66 86 POTENTIAL.  
CC TRANSMEM 93 113 POTENTIAL.  
CC DOMAIN 22 25 POLY-THR.  
CC DOMAIN 62 65 POLY-SER.  
CC DOMAIN 235 243 POLY-GLU.  
CC SEQUENCE 316 AA; 35046 MW; ADFAD73924279550 CRC64;  
CC -----  
CC Query Match 0.6%; Score 8; DB 1; Length 316;  
CC Best Local Similarity 100.0%; Pred. No. 12;  
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC 1325 EEEEEER 1332  
CC |  
CC 237 EEEEEER 244  
CC -----  
CC RESULT 12  
CC RLAA\_ARCFU STANDARD; PRT; 339 AA.  
CC ID RLAA\_ARCFU  
CC AC 028781;  
CC DT 15-DEC-1998 (Rel. 37, Created)  
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
CC DT 01-OCT-2000 (Rel. 40, Last annotation update)  
CC DE ACIDIC RIBOSOMAL PROTEIN P0 HOMOLOG (L10E).  
CC GN RPLP0 OR AFL491.  
CC OS Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
CC OC Archaeoglobus fulgidus.  
CC CC Archaeoglobus.  
CC NCBI\_TaxID=2234;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
CC MEDLINE=98049343; PubMed=9389475;  
CC Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
CC Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
CC Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
CC Raikess E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
CC Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
CC Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
CC Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
CC Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
CC Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
CC Venter J.C.;  
CC "The complete genome sequence of the hyperthermophilic, sulphate-  
CC reducing archaeon Archaeoglobus fulgidus";  
CC Nature 390:364-370(1997).  
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT  
CC OF E.COLI PROTEIN L10.  
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.



RC TISSUE-Eosinophil;  
RA MEDLINE-91160746; PubMed=1705901;  
RX Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M.,  
RF Saito T., Tasaka K.;  
RT "Sequencing and cloning of the cDNA of guinea pig eosinophil major  
RI basic protein.";  
RN FEBS Lett. 279:330-334 (1991).  
[2]  
RP PARTIAL SEQUENCE.  
RX MEDLINE-94092714; PubMed=8268206;  
RA Hashimoto Y., Nagaoka I., Yamashita T.;  
RF "Purification of the antibacterial fragments of guinea-pig major  
RI basic protein.";  
RN Biochim. Biophys. Acta 1203:236-242 (1993).  
[2]  
RP FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC  
RX REACTIONS AND INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING  
RA HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF  
RF BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.  
RI  
RN SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC  
RX GRANULE (CRYSTALLOID CORE).  
RA  
RF SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
RI  
RN This SWISS-PROT entry is copyright. It is produced through a collaboration  
RX between the Swiss Institute of Bioinformatics and the EMBL outstation -  
RA the European Bioinformatics Institute. There are no restrictions on its  
RF use by non-profit institutions as long as its content is in no way  
RI modified and this statement is not removed. Usage by and for commercial  
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
RX or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
RA  
RF EMBL; D90251; BAA14291.1; -  
RI PIR; S13625; S13625.  
RN InterPro; IPR001304; -  
RX InterPro; IPR002352; -  
RA Pfam; PF00059; Lectin\_G; 1.  
RF PRINTS; PR00770; EMABORBASCP.  
RI PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
RN PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
RX Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;  
RI Lectin; Multigene family.  
RN SIGNAL 1 15 POTENTIAL.  
RX PROPEP 16 114 ACIDIC.  
RI CHAIN 115 233 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.  
RN DOMAIN 132 233 C-TYPE LECTIN (SHORT FORM).  
RX DISULFID 134 231 BY SIMILARITY.  
RI DISULFID 208 223 BY SIMILARITY.  
RN SEQUENCE 233 AA; 26268 MW; C8D5E96D927C56C8 CRC64;  
[1]  
Query Match 0.6%; Score 8; DB 1; Length 233;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1324 PEEEEEE 1331  
Db 73 PEEEEEE 80  
RESULT 9  
SQM1\_ONCKE STANDARD; PRT; 233 AA.  
AC P24405;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE SOMATOLACTIN PRECURSOR (SL).  
OS Oncorhynchus keta (Chum salmon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8018;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX

RX MEDLINE-92017883; PubMed=1922096;  
RA Takayama Y., Rand-Weaver M., Kawauchi H., Ono M.;  
RT "Gene structure of chum salmon somatolactin, a presumed pituitary  
RF hormone of the growth hormone/prolactin family.";  
RI Mol. Endocrinol. 5:778-786 (1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92038879; PubMed=1936917;  
RA Takayama Y., Ono M., Rand-Weaver M., Kawauchi H.;  
RF "Greater conservation of somatolactin, a presumed pituitary hormone  
RI of the growth hormone/prolactin family, than of growth hormone in  
RN teleost fish.";  
RP Gen. Comp. Endocrinol. 83:366-374 (1991).  
RX FUNCTION: MAY BE ASSOCIATED WITH ION REGULATION AND REPRODUCTION.  
RA SUBCELLULAR LOCATION: SECRETED.  
RF TISSUE SPECIFICITY: PITUITARY GLAND.  
RI SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
RN  
RX This SWISS-PROT entry is copyright. It is produced through a collaboration  
RA between the Swiss Institute of Bioinformatics and the EMBL outstation -  
RF the European Bioinformatics Institute. There are no restrictions on its  
RI use by non-profit institutions as long as its content is in no way  
RN modified and this statement is not removed. Usage by and for commercial  
RX entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
RA or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
RF EMBL; D10640; BAA01487.1; -  
RI EMBL; D10638; BAA01485.1; -  
RN EMBL; D10636; BAA01485.1; JOINED.  
RX EMBL; D10637; BAA01485.1; JOINED.  
RI PIR; A23729; A23729.  
RN PIR; B49762; B49762.  
RX HSSP; P01241; 1BP3.  
RI InterPro; IPR001400; -  
RN Pfam; PF00103; hormone; 1.  
RX PRINTS; PR00836; SOMATOTROPIN.  
RI PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
RN PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
RX Hormone; Signal.  
RI SIGNAL 1 24 POTENTIAL.  
RN CHAIN 25 233 SOMATOLACTIN.  
RX DISULFID 29 39 BY SIMILARITY.  
RI DISULFID 89 205 BY SIMILARITY.  
RN DISULFID 222 230 BY SIMILARITY.  
RX SEQUENCE 233 AA; 26882 MW; 178F385705F0A461 CRC64;  
[1]  
Query Match 0.6%; Score 8; DB 1; Length 233;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 979 VSEKLSSL 986  
Db 150 VSEKLSSL 157  
RESULT 10  
THYL\_HUMAN STANDARD; PRT; 242 AA.  
ID AC P20396;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE THYROLIBERIN PRECURSOR (THYROTROPIN RELEASING HORMONE) (TRH).  
GN TRH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91125361; PubMed=2126343;  
RA Yamada M., Radovick S., Wondisford F.E., Nakayama Y., Weintraub B.D.,



KW	Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW	Oxidoreductase; Transit peptide.
FT	1 13 MITOCHONDRION.
FT	14 91 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT	17 25 11 KDA PROTEIN.
FT	DOMAIN
FT	SEQUENCE 91 AA; 10755 MW; 499844E100138E4C CRC64;
SQ	

Query Match 0.6%; Score 8; DB 1; Length 91;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6	
VE4_HPVO4	
ID	VE4_HPVO4
STANDARD;	PRT; 181 AA.
AC	Q07852;
DT	01-OCT-1994 (Rel. 30, Created)
DT	01-OCT-1994 (Rel. 30, Last sequence update)
DT	01-OCT-1994 (Rel. 30, Last annotation update)
DE	PROBABLE E4 PROTEIN.
CS	Human papillomavirus type 4.
OS	Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
OX	NCBI TaxID:10617;

```
Query Match      0.6%; Score 8; DB 1; Length 181;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT	7
HMG2_CHICK	
ID	HMG2_CHICK STANDARD; PRT; 206 AA.
AC	P26584;
DT	01-AUG-1992 (Rel. 23, Created)
DI	01-AUG-1992 (Rel. 23, Last sequence update)
DE	01-FEB-1995 (REL. 31, Last annotation update)
DE	HIGH MOBILITY GROUP PROTEIN HMG2 (HMG-2).
GN	HMG2.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

OC Gallus.
NCBI_TaxID=9031;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92241676; PubMed=1572546;
RA Davis D.L., Burch J.B.E.;
RT "Isolation of a chicken HMG2 cDNA clone and evidence for an HMG2-
RL specific 3'-untranslated region.";
RN Gene 113:251-256(1992).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92290291; PubMed=1601311;
RA Sparrow D.B., Wells J.R.E.;
RT "Sequence of a cDNA encoding chicken high-mobility-group protein-2." ;
RN Gene 114:289-290(1992).
CC -!- FUNCTION: BINDS PREFERENTIALLY SINGLE-STRANDED DNA AND UNWINDS
CC DOUBLE STRANDED DNA.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HMGI/HMG2 PROTEIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 HMG BOXES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensel@lsb-sib.ch).
CC -----

```

RESULT	8	
EMBL_CAVPO		
ID	EMBL_CAVPO	STANDARD;
AC	P22032	
DT	01-AUG-1991	(Rel. 19, Created)
DT	01-AUG-1991	(Rel. 19, Last sequence update)
DT	15-DEC-1998	(Rel. 37, Last annotation update)
DE	DE FOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1	PRECURSOR (MBP-1).
GN	MBP1	
OS	Cavia porcellus	(Guinea pig).
OC	Eukaryota; Metazoa; Chordata;	Cranial; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia;	Hystricognathi; Caviidae; Cavia.
NCBI_TaxID	10141	
RP	[1]	
RP	SEQUENCE FROM N.A.,	AND SEQUENCE OF 115-162.



CC modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z15112; CAA78816.1; -  
CC PIR: S25626; S25626.  
CC PIR: S17596; S17596.  
CC PIR: S27275; S27275.  
CC InterPro: IPR001770;  
CC PROSITE: PS00058; G-PROTEIN GAMMA; FALSE NEG.  
KW Transducer; Prenylation; Lipoprotein; Vision.  
FT DOMAIN 7 11 POLY-LYS.  
FT MOD\_RES 12 18 POLY-GLU.  
FT LIPID 84 84 BLOCKED.  
FT PROPEP 85 87 GERANYL-GERANYL (BY SIMILARITY).  
FT PROPEP 85 87 REMOVED IN MATURE FORM (BY SIMILARITY).  
SQ SEQUENCE 87 AA; 10278 MW; BF9E07DEB780B104 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 87;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
DB 12 EEEEEER 19

RESULT 5  
UCRH\_HUMAN  
ID UCRH\_HUMAN STANDARD; PRT; 91 AA.  
AC P07919;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KDA PROTEIN PRECURSOR  
DE (EC 1.10.2.2) (MITOCHONDRIAL HINGE PROTEIN) (CYTOCHROME C1, NONHEME 11  
DE KDA PROTEIN) (COMPLEX III SUBUNIT VIII).  
GN UQCRH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88083627; PubMed=2826252;  
RA Ohta S., Goto K., Arai H., Kagawa Y.;  
RT "An extremely acidic amino-terminal presequence of the precursor for  
the human mitochondrial hinge protein.";  
RL FEBS Lett. 226:171-175(1987).  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C  
REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH  
IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY  
MEDIATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.  
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C -> Q +  
2 FERROCYTOCHROME C.  
CC -!- SUBUNIT: BCI COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS  
(CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND  
6 LOW-MOLECULAR WEIGHT PROTEINS.  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE UQCRH/OCR6 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M36647; AAA36317.1; -  
CC EMBL: Y00764; CAA68733.1; -  
CC PIR: S00219; S00219.

CC modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z15112; CAA78816.1; -  
CC PIR: S25626; S25626.  
CC PIR: S17596; S17596.  
CC PIR: S27275; S27275.  
CC InterPro: IPR001770;  
CC PROSITE: PS00058; G-PROTEIN GAMMA; FALSE NEG.  
KW Transducer; Prenylation; Lipoprotein; Vision.  
FT DOMAIN 7 11 POLY-LYS.  
FT MOD\_RES 12 18 POLY-GLU.  
FT LIPID 84 84 BLOCKED.  
FT PROPEP 85 87 GERANYL-GERANYL (BY SIMILARITY).  
FT PROPEP 85 87 REMOVED IN MATURE FORM (BY SIMILARITY).  
SQ SEQUENCE 87 AA; 10278 MW; BF9E07DEB780B104 CRC64;

Query Match 0.6%; Score 9; DB 1; Length 1938;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKQLEV 565  
DB 1528 EKIRKQLEV 1536

RESULT 4  
GBG\_LOLFO  
ID GBG\_LOLFO STANDARD; PRT; 87 AA.  
AC Q01821.  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE GUANINE NUCLEOTIDE-BINDING PROTEIN GAMMA SUBUNIT.  
DE Loligo forbesi (Northern European squid).  
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;  
OC Myopsida; Loliginidae; Loligo.  
OX NCBI\_TaxID=6618;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX TISSUE=Retina;  
RA Lott J.S., Ryba N.J.P., Pottinger J.D.D., Keen J.N., Carne A.,  
Findlay J.B.C.;  
RT "The gamma-subunit of the principal G-protein from squid (Loligo  
forbesi) photoreceptors contains a novel N-terminal sequence.";  
RL FEBS Lett. 312:241-244(1992).  
CC [2]  
CC PRELIMINARY PARTIAL SEQUENCE OF 28-72.  
CC MEDLINE=92028830; PubMed=1930153;  
CC Pottinger J.D.D., Ryba N.J.P., Keen J.N., Findlay J.B.C.;  
CC "The identification and purification of the heterotrimeric  
GTP-binding protein from squid (Loligo forbesi) photoreceptors.";  
CC Biochem. J. 279:323-326(1991).  
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
SIGNALING SYSTEMS. THIS MAJOR G-PROTEIN OF THE SQUID PHOTORECEPTOR  
IS INVOLVED IN VISUAL TRANSDUCTION. THE BETA AND GAMMA CHAINS ARE  
REQUIRED FOR THE GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP,  
AND FOR G PROTEIN-EFFECTOR INTERACTION.  
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
CC -!- SIMILARITY: BELONGS TO THE G PROTEIN GAMMA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way



EMBL; M76599; AAA37160.1; -  
 EMBL; M76600; AAA37161.1; -  
 EMBL; M76601; AAA37162.1; -  
 EMBL; M62404; AAA37424.1; -  
 HSSP; P08799; 1MND.  
 SWISS-2DPAGE; Q02566; MOUSE.  
 MGD; MGI:97255; Myhca.  
 InterPro; IPR000048; -  
 InterPro; IPR001609; -  
 InterPro; IPR002928; -  
 Pfam; PF00612; IQ; 1.  
 Pfam; PF01576; Myosin\_tail; 1.  
 Pfam; PF00063; myosin\_head; 1.  
 PRINTS; PR00193; MYOSINHEAVY.  
 PROSITE; PS50096; IQ; 1.  
 Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 ATP-binding; Methylation; Alkylation; Heptad repeat pattern;  
 Multigene family; Calmodulin-binding; Polymorphism.  
 DOMAIN 1 841  
 GLOBULAR HEAD (S1).  
 DOMAIN 842 1938  
 RODLIKE TAIL (S2 AND LMM DOMAINS).  
 NP\_BIND 178 185  
 COILED COIL (POTENTIAL).  
 DOMAIN 657 679  
 ACTIN-BINDING.  
 DOMAIN 759 773  
 ACTIN-BINDING.  
 DOMAIN 790 807  
 CALMODULIN-BINDING (BY SIMILARITY).  
 DOMAIN 816 833  
 CALMODULIN-BINDING (BY SIMILARITY).  
 MOD\_RES 129 129  
 METHYLATION (TRP-) (POTENTIAL).  
 MOD\_RES 697 697  
 ALKYLATION (SH-1) (POTENTIAL).  
 MOD\_RES 707 707  
 ALKYLATION (SH-2) (POTENTIAL).  
 VARIANT 194 194  
 Y -> D.  
 VARIANT 545 545  
 S -> A.  
 VARIANT 838 838  
 I -> S.  
 SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 0.6%; Score 9; DB 1; Length 1938;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

557 EKIRKQLEV 565  
 |||||  
 1529 EKIRKQLEV 1537

RESULT 3  
 MESA\_RAT STANDARD; PRT; 1938 AA.  
 AC P02563; Q63351;  
 DE 21-JUL-1986 (Rel. 01, Created)  
 DE 01-OCT-1989 (Rel. 12, Last sequence update)  
 DE 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.  
 GN MYH6.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 [1]  
 SEQUENCE FROM N.A.  
 TISSUE=Heart;  
 MEDLINE=90016822; PubMed=2798111;  
 Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
 "Complete nucleotide sequence of full length cDNA for rat alpha  
 cardiac myosin heavy chain."  
 Nucleic Acids Res. 17:7527-7528(1989).  
 [2]  
 DISCUSSION OF SEQUENCE.  
 MEDLINE=90133919; PubMed=2614840;  
 McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
 "Full-length rat alpha and beta cardiac myosin heavy chain sequences.  
 Comparisons suggest a molecular basis for functional differences."  
 J. Mol. Biol. 210:665-671(1989).  
 [3]

RP SEQUENCE OF 1-167 FROM N.A.  
 RX MEDLINE=84194059; PubMed=6585819;  
 RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;  
 RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in  
 tandem.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).  
 RN [4]  
 RP SEQUENCE OF 1512-1938 FROM N.A.  
 RX MEDLINE=82220036; PubMed=7045682;  
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;  
 RT "Molecular characterization of two myosin heavy chain genes expressed  
 in the adult heart."  
 RL Nature 297:659-664(1982).  
 RN [5]  
 RP SEQUENCE OF 1872-1938 FROM N.A.  
 RC STRAIN=Wistar; Tissue=Heart;  
 RX MEDLINE=85179510; PubMed=6241892;  
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;  
 RT "Cardiac myosin heavy chain isozymic transitions during development  
 and under pathological conditions are regulated at the level of mRNA  
 availability.";  
 RL Eur. Heart J. 5:181-191(1984).  
 CC -!- FUNCTION: MUSCLE CONTRACTION.  
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE  
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 SUBFRAGMENT (S2).  
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE  
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.  
 CC -!- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN  
 THE RAT, TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY  
 CHAINS.  
 CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,  
 WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE  
 CONSERVED.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X15938; CAA34064.1; -  
 EMBL; K01464; AAA41648.1; -  
 EMBL; J00751; AAA41653.1; -  
 EMBL; M32697; AAA41658.1; -  
 PIR; S06005; S06005.  
 PIR; A02988; A02988.  
 HSSP; P08799; 1MND.  
 InterPro; IPR000048; -  
 InterPro; IPR001609; -  
 InterPro; IPR002928; -  
 Pfam; PF00612; IQ; 1.  
 Pfam; PF01576; Myosin\_tail; 1.  
 Pfam; PF00063; myosin\_head; 1.  
 PRINTS; PR00193; MYOSINHEAVY.  
 PROSITE; PS50096; IQ; 1.  
 Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 ATP-binding; Methylation; Alkylation; Heptad repeat pattern;  
 Multigene family; Calmodulin-binding;  
 DOMAIN 1 840  
 GLOBULAR HEAD (S1).  
 DOMAIN 841 1938  
 RODLIKE TAIL (S2 AND LMM DOMAINS).  
 FT DOMAIN



```

983 0.4 312 1 DBX-ANAPL O57314 anas platyr
984 0.4 312 1 YR81-CAEEL Q09566 caenorhabdi
985 0.4 313 1 N1M1-BRALA Q21000 branchiost
986 0.4 313 1 YDJB-SCHPO P87059 schizosacch
987 0.4 314 1 ACCO-MALDO Q00985 malus domes
988 0.4 314 1 LEUO-ECOLI P10151 escherichia
989 0.4 316 1 DHPS-STRPN P05382 streptococc
990 0.4 316 1 YDEB-BACSU Q06987 bacillus su
991 0.4 317 1 AAC4-DICDI P14198 dictyosteli
992 0.4 317 1 AC28-CANAL P43063 candida alb
993 0.4 318 1 CYPR-YEAST P25334 saccharomyc
994 0.4 318 1 Y051-NPVO P10309 orgyia pseu
995 0.4 318 1 Y211-AQUAE O68405 aquifex aeo
996 0.4 318 1 Y220-AQUAE O68411 aquifex aeo
997 0.4 319 1 YMP-CERV P05396 carnation e
998 0.4 321 1 IPNS-STRCT Q53932 streptomyce
999 0.4 321 1 Y4JO-RHISN P55515 rhizobium s
1000 0.4 322 1 SUCA-CAEEL P53596 caenorhabdi

ALIGNMENTS
TRFL_HUMAN STANDARD; PRT: 439 AA.
AC P54274;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE TELOMERIC REPEAT BINDING FACTOR 1.
GN TRFL OR TRF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RE SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96099400; PubMed=7502076;
RA Chong L., van Steensel B., Broccoli D., Erdjument-Bromage H.,
RA Hanish J., Tempst P., de Lange T.;
RT "A human telomeric protein.";
RT Science 270:1663-1667(1995).
RN [2]
RE REVISION TO 14.
RA de Lange T.;
RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RE STRUCTURE BY NMR OF 378-430.
RX MEDLINE=98416700; PubMed=9739097;
RA Nishikawa T., Nagadol A., Yoshimura S., Almoto S., Nishimura Y.;
RT "Solution structure of the DNA-binding domain of human telomeric
RT protein, hTRF1.";
RT Structure 6:1057-1065(1998).
CC -!- FUNCTION: BINDS THE TELOMERIC DOUBLE-STRANDED TTAGGG REPEAT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZES WITH TELOMERIC DNA IN
CC METAPHASE CELLS AND IS LOCATED AT CHROMOSOME ENDS DURING
CC METAPHASE.
CC -!- SIMILARITY: CONTAINS A MYB-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40705; -; NOT_ANNOTATED_CDS.
CC PDB; 1BA5; 27-APR-99.
CC MIM; 600951; -;
CC PROSITE; PS50090; MYB_3; 1.
CC DNA-binding; Telomere; Nuclear protein; 3D-structure.

```

```

FT DOMAIN 55 62 POLY-GLU.
FT DOMAIN 349 352 POLY-LYS.
FT DNA_BIND 375 428 MYB.
SQ SEQUENCE 439 AA: 50344 MW: 44FA3E7FF3E9CCAE CRC64;

Query Match 0.6%; Score 9; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331
DB 53 APEEEEEEE 61

RESULT 2
MISA_MOUSE STANDARD; PRT: 1938 AA.
AC Q02566; Q64258; Q64738;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.
GN MYH6 OR MYHCA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RX STRAIN=BALE/C, A/J, C57BL/6J, AND DBA/2J;
RX MEDLINE=92250040; PubMed=1577481;
RA Quinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;
RT "Characterization of the allelic differences in the mouse cardiac
RT alpha-myosin heavy chain coding sequence.";
RT Genomics 13:176-188(1992).
RN [2]
RE SEQUENCE OF 1-67 FROM N.A.
RX STRAIN=AKR;
RX MEDLINE=91225025; PubMed=2026617;
RA Gulick J., Subramaniam A., Neumann J., Robbins J.;
RT "Isolation and characterization of the mouse cardiac myosin heavy
RT chain genes.";
RT J. Biol. Chem. 266:9180-9185(1991).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M76598; AAA37159.1; -;

```



837	1	PPA_ZYMMO	264	0.4	6	P14924 zymomonas m	910	0.4	291	1	CD20_MOUSE	P19437 mus musculus
838	1	TRPC_ZYMP	265	0.4	6	Q9xbm3 zymomonas m	911	0.4	291	1	CGD2_XENLA	P53782 xenopus lae
839	1	YCFH_ECOLI	265	0.4	6	P37346 escherichia	912	0.4	291	1	FLAV_MATIN	O04395 mathiola i
840	1	DHSE_YEAST	266	0.4	6	P21801 saccharomyc	913	0.4	292	1	AMPF_PROST	O69772 providencia
841	1	NIFW_KLEPN	266	0.4	6	P08534 klebsiella	914	0.4	292	1	BIP2_TOBAC	Q03682 nicotiana t
842	1	YD37_SCHPO	266	0.4	6	Q10271 schizosacch	915	0.4	292	1	CC22_ORYSA	P29619 oryza sativ
843	1	143H_ARATH	267	0.4	6	P46077 arabidopsis	916	0.4	292	1	CC2H_DICDI	P34117 dictyosteli
844	1	CGIC_DROME	267	0.4	6	P25008 drosophila	917	0.4	293	1	TYRC_ZYMMO	Q04983 zymomonas m
845	1	PENK_HUMAN	267	0.4	6	P01210 homo sapien	918	0.4	294	1	ARRL_YEAST	Q06596 saccharomyc
846	1	PKI_NPVHZ	267	0.4	6	P41719 heliothis z	919	0.4	294	1	CC21_ORYSA	P29618 oryza sativ
847	1	SUHB_HAEIN	267	0.4	6	P44333 haemophilus	920	0.4	294	1	CC22_MEDSA	Q05006 medicago sa
848	1	CHHL_HUMAN	268	0.4	6	P53701 homo sapien	921	0.4	294	1	CC2A_ANTHA	Q38772 antirrhinum
849	1	ILIA_BOVIN	268	0.4	6	P08831 bos taurus	922	0.4	294	1	CC2A_ARATH	P24100 arabidopsis
850	1	ILIA_CAPHI	268	0.4	6	P79161 capra hircu	923	0.4	294	1	CC2_MAIZE	P23111 zea mays (m
851	1	ILIA_SHEEP	268	0.4	6	Q28579 ovis aries	924	0.4	294	1	CC2_VIGUN	P52389 vigna ungu
852	1	INHA_MYCAV	268	0.4	6	Q07400 mycobacteri	925	0.4	294	1	SMN_HUMAN	Q16637 homo sapien
853	1	ISPE_HELPY	268	0.4	6	O25984 helicobacte	926	0.4	295	1	CPPM_SALTY	Q05062 salmonella
854	1	RAD_RAT	268	0.4	6	P55043 rattus norv	927	0.4	295	1	MAUJ_METEX	Q49127 methylobact
855	1	TRPA_ECOLI	268	0.4	6	P00928 escherichia	928	0.4	296	1	CPPM_ECOLI	P77541 escherichia
856	1	TRPA_SALTY	268	0.4	6	P00929 salmonella	929	0.4	296	1	EGR3_MOUSE	P43300 mus musculus
857	1	VANY_ENTFA	268	0.4	6	Q47746 enterococcu	930	0.4	296	1	THIM_RAT	P97532 rattus norv
858	1	RAD_HUMAN	269	0.4	6	P55042 homo sapien	931	0.4	296	1	YBNB_ECOLI	P77395 escherichia
859	1	TRPA_KLEAE	269	0.4	6	P00930 klebsiella	932	0.4	297	1	CD20_HUMAN	P11836 homo sapien
860	1	VC03_SPKVA	269	0.4	6	P32229 swinepox vi	933	0.4	297	1	DSR3_MOUSE	O14972 homo sapien
861	1	VG51_HSVSA	269	0.4	6	Q01036 herpesvirus	934	0.4	297	1	DSR3_MOUSE	Q35075 mus musculus
862	1	CDSA_BRUAB	270	0.4	6	O59173 b phosphati	935	0.4	297	1	MYOD_COTJA	P21572 coturnix co
863	1	ILIA_HORSE	270	0.4	6	Q28385 equus cabal	936	0.4	297	1	RRPP_RABVA	P15398 rabies viru
864	1	ILIA_MOUSE	270	0.4	6	P01582 mus musculus	937	0.4	297	1	RRPP_RABVC	P22363 rabies viru
865	1	PROC_CORGL	271	0.4	6	P46540 corynebacte	938	0.4	297	1	RRPP_RABVE	P22559 rabies viru
866	1	TRPA_MYCIT	271	0.4	6	O68906 mycobacteri	939	0.4	297	1	RRPP_RABVP	P06747 rabies viru
867	1	YAHF_ECOLI	271	0.4	6	P77393 escherichia	940	0.4	297	1	RRPP_RABVS	P16286 rabies viru
868	1	VG5W_YEAST	271	0.4	6	P53335 saccharomyc	941	0.4	297	1	VGIG_HRSVA	P27023 human respi
869	1	YPOQ_KLEPN	271	0.4	6	P27509 klebsiella	942	0.4	297	1	VGIG_HRSV6	P27025 human respi
870	1	DHKL_STRVN	272	0.4	6	P16542 streptomyce	943	0.4	297	1	VGIG_HRSV7	P27026 human respi
871	1	VALI_MSVN	272	0.4	6	P03568 maize strea	944	0.4	298	1	CC28_YEAST	P00546 saccharomyc
872	1	VALI_MSVN	272	0.4	6	P14980 maize strea	945	0.4	298	1	MMSB_PSEAE	P28811 pseudomonas
873	1	VALI_MSVN	272	0.4	6	P14990 maize strea	946	0.4	298	1	MYOD_CHICK	P16075 gallus gall
874	1	OSAL_BORBU	273	0.4	6	P14013 borrelia bu	947	0.4	298	1	VGIG_HRSV5	P27024 human respi
875	1	T2C2_HERAU	273	0.4	6	P25259 herpetosiph	948	0.4	298	1	VGIG_HRSVA	P03423 human respi
876	1	TCIA_CAEBR	273	0.4	6	P35072 caenorhabdi	949	0.4	298	1	VGIG_HRSVL	P20895 human respi
877	1	TC2A_CAEBR	273	0.4	6	Q04202 caenorhabdi	950	0.4	299	1	YF02_MYCTU	P71783 mycobacteri
878	1	PSBS_SPIOL	274	0.4	6	Q02060 spinacia ol	951	0.4	299	1	YF02_MYCTU	P37513 bacillus su
879	1	AMOA_NITEU	275	0.4	6	Q04507 nitrosomona	952	0.4	300	1	CPPM_BACSD	Q92977 bacillus ha
880	1	CDID_SYLEL	275	0.4	6	P23043 sylvilagus	953	0.4	300	1	NIFU_ANAHD	Q43885 anabaena az
881	1	LEC_PEA	275	0.4	6	P02867 pisum sativ	954	0.4	300	1	SPRC_XENLA	P36378 xenopus lae
882	1	YQUG_BACSU	275	0.4	6	P54544 bacillus su	955	0.4	300	1	T2D3_HAEIN	P43870 haemophilus
883	1	DLDH_RHET	277	0.4	6	O05940 rhizobium e	956	0.4	301	1	KIPN_BPT4	P06855 bacterioph
884	1	HAXI_HUMAN	279	0.4	6	O00165 homo sapien	957	0.4	301	1	MCP_BPF41	P26596 lactococcus
885	1	THET_THEVU	279	0.4	6	P04072 thermoactin	958	0.4	301	1	OPSD_CAMSC	O16018 cambarellus
886	1	AROE_BACSU	280	0.4	6	P54374 bacillus su	959	0.4	301	1	SCRK_ZYMMO	Q03417 zymomonas m
887	1	CC2B_ANTMA	280	0.4	6	Q38773 antirrhinum	960	0.4	301	1	TRP2_CHICK	P02642 gallus gall
888	1	DNAI_THETH	280	0.4	6	Q56237 thermus aqu	961	0.4	302	1	FABD_MYCTU	Q10501 mycobacteri
889	1	RP32_HAEIN	281	0.4	6	P44404 haemophilus	962	0.4	306	1	DHB3_RAT	Q54939 rattus norv
890	1	RP32_METJA	281	0.4	6	Q60180 methanococc	963	0.4	306	1	MPRI_YEAST	P43588 saccharomyc
891	1	YTXC_BACSU	281	0.4	6	P06569 bacillus su	964	0.4	306	1	OPPB_ECOLI	P31132 escherichia
892	1	PANC_CAMJE	282	0.4	6	Q9pik2 campylobact	965	0.4	306	1	OPPB_SALTY	P08005 salmonella
893	1	FLAB_LEPIN	283	0.4	6	O51941 leptospira	966	0.4	306	1	YBFE_BACSU	O31448 bacillus su
894	1	RP38_HUMAN	283	0.4	6	P78345 homo sapien	967	0.4	306	1	IDFC_BACSU	P96880 bacillus su
895	1	RP32_PROMI	284	0.4	6	P50509 proteus mir	968	0.4	307	1	SRB5_YEAST	P32585 saccharomyc
896	1	YURL_BACSU	284	0.4	6	O32153 bacillus su	969	0.4	308	1	FRP_HUMAN	Q12841 homo sapien
897	1	RP32_ENTCL	285	0.4	6	P50508 enterobacte	970	0.4	308	1	KHSE_BRELA	P07128 revibacter
898	1	YAT1_SCHPO	285	0.4	6	Q10148 schizosacch	971	0.4	308	1	KHSE_CORGL	P08210 corynebacte
899	1	YD98_YEAST	285	0.4	6	Q03835 saccharomyc	972	0.4	308	1	MENA_HAEIN	P44739 haemophilus
900	1	FOLD_SALTI	287	0.4	6	Q60006 salmonella	973	0.4	308	1	YBIN_ECOLI	P75782 escherichia
901	1	MF25_XENLA	287	0.4	6	P16076 xenopus lae	974	0.4	310	1	E13A_HORVU	P34742 hordeum vul
902	1	DSR2_HUMAN	288	0.4	6	O95456 homo sapien	975	0.4	310	1	GTRB_BPP22	P57022 bacterioph
903	1	ILVE_METJA	288	0.4	6	Q58414 methanococc	976	0.4	310	1	HEM2_BACHD	O9kdk9 bacillus ha
904	1	LIPH_PSEAE	288	0.4	6	Q01725 pseudomonas	977	0.4	310	1	PURU_MYCTU	O50453 mycobacteri
905	1	CAH8_HUMAN	290	0.4	6	P35219 homo sapien	978	0.4	310	1	VU84_HSV7J	P52534 human herpe
906	1	GSPC_AERSA	290	0.4	6	P45772 aeromonas s	979	0.4	310	1	Y4RE_RHISN	P55638 rhizobium s
907	1	TRX2_MOUSE	290	0.4	6	O08550 mus musculus	980	0.4	310	1	Y4RE_RHISN	P03207 epstein-bar
908	1	ATPG_BUCAP	291	0.4	6	O51873 buchnera ap	981	0.4	311	1	ATH5_ARATH	P46668 arabidopsis
909	1	CC21_MEDSA	291	0.4	6	P24923 medicago sa	982	0.4	312	1	ATH5_ARATH	P46667 arabidopsis



691	1	196	0.4	6	028220	archaeoglob	764	233	1	RR2_CYACA	P35014	cyaniidum c
692	1	196	0.4	6	P38472	marchantia	765	233	1	YB63_METJA	Q58858	methanococc
693	1	197	0.4	6	P05422	phyllomedus	766	234	1	MAG8_HUMAN	P43361	homo sapien
694	1	198	0.4	6	P05421	phyllomedus	767	234	1	SAMP_MESAU	P07629	mesocricetu
695	1	199	0.4	6	O32481	legionella	768	234	1	SNG1_RAT	Q62876	rattus norv
696	1	199	0.4	6	O05386	bacillus ce	769	234	1	YJ72_YEAST	P39543	saccharomyc
697	1	199	0.4	6	O84971	rhodobacter	770	235	1	END1_ECOLI	P25736	escherichia
698	1	201	0.4	6	P00106	pseudomonas	771	235	1	FLIH_SALTY	P15934	salmonella
699	1	201	0.4	6	P27344	saccharomyc	772	235	1	GAMT_RAT	P10868	rattus norv
700	1	202	0.4	6	O04905	arabidopsis	773	235	1	MTN_BUCAP	O51931	buchnera ap
701	1	203	0.4	6	P15771	homo sapien	774	235	1	PYG_AGLNE	P34814	aglaothamni
702	1	203	0.4	6	P53288	saccharomyc	775	235	1	TFE2_ALCEU	P94136	alcaligenes
703	1	204	0.4	6	P45273	haemophilus	776	235	1	VINC_XENLA	Q04615	xenopus lae
704	1	204	0.4	6	Q05163	dicertrarch	777	236	1	CYSA_CHLVU	P56344	chlorella v
705	1	204	0.4	6	Q01283	lates calca	778	236	1	GAMT_HUMAN	Q14353	homo sapien
706	1	204	0.4	6	P48248	morone saxa	779	236	1	GAMT_MOUSE	O35969	mus musculus
707	1	204	0.4	6	P09539	seriola qui	780	236	1	PLC1_BOVIN	P09611	bos taurus
708	1	204	0.4	6	P09113	thunus thy	781	237	1	NDF3_HUMAN	O92886	homo sapien
709	1	204	0.4	6	Q16890	homo sapien	782	238	1	CHVI_RHISN	P50351	rhizobium s
710	1	204	0.4	6	O54818	mus musculus	783	239	1	DCUR_ECOLI	P39271	escherichia
711	1	204	0.4	6	Q06600	bos taurus	784	239	1	UPPS_HAEIN	P44938	haemophilus
712	1	205	0.4	6	P34555	caenorhabdi	785	240	1	CHVI_RHIME	P50350	rhizobium m
713	1	206	0.4	6	P05205	drosophila	786	240	1	HDFE_HUMAN	P51858	homo sapien
714	1	206	0.4	6	Q00630	manduca sex	787	240	1	PYRH_BACSU	O31749	bacillus su
715	1	206	0.4	6	Q58298	methanococc	788	241	1	AGL8_SINAL	Q41274	sinapis alb
716	1	206	0.4	6	P76264	escherichia	789	241	1	CHVI_AGRTO	Q07783	agrobacteri
717	1	208	0.4	6	P28074	homo sapien	790	241	1	PHNF_ECOLI	P16684	escherichia
718	1	208	0.4	6	Q12488	saccharomyc	791	242	1	AGL8_ARATH	Q38876	arabidopsis
719	1	209	0.4	6	Q04603	human immu	792	242	1	ARTJ_ECOLI	Q30860	escherichia
720	1	209	0.4	6	O51560	borrelia bu	793	243	1	TONB_ENTAE	P46383	enterobacte
721	1	210	0.4	6	P27682	rattus norv	794	243	1	TPX1_HUMAN	P16562	homo sapien
722	1	210	0.4	6	Q91814	gallus gall	795	243	1	TRIC_XENLA	P50754	xenopus lae
723	1	210	0.4	6	P51560	vibrio angu	796	244	1	SANA_HAEIN	P45130	haemophilus
724	1	210	0.4	6	P75256	mycoplasma	797	244	1	YKL4_CAEEL	P42171	caenorhabdi
725	1	210	0.4	6	Q11071	caenorhabdi	798	245	1	CRS3_HORSE	Q19010	equus cabal
726	1	211	0.4	6	P26487	azorhizobiu	799	245	1	NIS1_LACLA	P42708	lactococcus
727	1	211	0.4	6	O08849	mus musculus	800	245	1	TM4B_HUMAN	Q94918	homo sapien
728	1	211	0.4	6	P33407	oryza sativ	801	245	1	YBPA_BURCE	P37335	burkholderi
729	1	211	0.4	6	P34452	caenorhabdi	802	246	1	CHID_LYCES	Q05537	lycopersico
730	1	211	0.4	6	Q04550	southampton	803	246	1	INTL_ECOLI	Q47036	escherichia
731	1	211	0.4	6	P45390	escherichia	804	246	1	INTM_ECOLI	P56979	escherichia
732	1	212	0.4	6	P12961	mus musculus	805	246	1	MYP0_HETFR	P20938	heterodontu
733	1	212	0.4	6	P09457	saccharomyc	806	246	1	YBEC_CLOHI	Q92940	clostridium
734	1	212	0.4	6	Q02130	lactococcus	807	247	1	GRPB_RAT	P08462	rattus norv
735	1	213	0.4	6	P29227	drosophila	808	247	1	DJ_DRONE	P22588	bacterioph
736	1	213	0.4	6	O94468	schizosacch	809	248	1	VGS_BPPI	O01352	drosophila
737	1	214	0.4	6	P14204	bacillus su	810	250	1	AMPM_MYCGA	Q52353	mycoplasma
738	1	214	0.4	6	P92847	bothriopsis	811	252	1	AGSD_ECOLI	P46004	escherichia
739	1	214	0.4	6	P22378	simian immu	812	252	1	DHSB_SCHPO	P21911	schizosacch
740	1	215	0.4	6	Q00717	emericeella	813	252	1	DUT_HUMAN	P33316	homo sapien
741	1	215	0.4	6	P75849	escherichia	814	252	1	NUOE_MYCTU	P95177	mycobacteri
742	1	216	0.4	6	P77985	staphylococ	815	252	1	PRGK_SALTY	P41786	salmonella
743	1	217	0.4	6	O60046	neurospora	816	254	1	COBJ_PSEDE	P21640	pseudomonas
744	1	217	0.4	6	P28161	homo sapien	817	255	1	ISPE_CAMUE	Q9np10	campylobact
745	1	217	0.4	6	P09629	homo sapien	818	255	1	SKGA_CAUCR	Q9rp67	caulobacter
746	1	217	0.4	6	P09024	mus musculus	819	256	1	ADH_DROTS	P31550	drosophila
747	1	217	0.4	6	P54601	bacillus su	820	256	1	OBFI_MOUSE	Q64693	mus musculus
748	1	218	0.4	6	P91580	ciona intes	821	257	1	MSRA_BRANA	P54151	brassica na
749	1	219	0.4	6	P54536	bacillus su	822	257	1	PTC1_CYNPY	O42353	cynops pyrr
750	1	220	0.4	6	Q91771	xenopus lae	823	258	1	MSRA_ARATH	P54150	arabidopsis
751	1	220	0.4	6	P37502	bacillus su	824	258	1	ULJ4_HSV7J	P52466	human herpe
752	1	221	0.4	6	P18432	drosophila	825	258	1	Y875_SYNY3	P73555	synecocyst
753	1	221	0.4	6	P19972	pichia fari	826	258	1	YO33_BPHP1	P51737	bacterioph
754	1	225	0.4	6	P33999	escherichia	827	259	1	MSRA_LACSA	Q9sec2	lactuca sat
755	1	226	0.4	6	P78035	mycoplasma	828	260	1	CABV_MOUSE	P12658	mus musculus
756	1	226	0.4	6	P04978	bordetella	829	261	1	MURI_BORBU	O51127	borrelia bu
757	1	226	0.4	6	P42399	bacillus su	830	261	1	SPRE_HUMAN	P35270	homo sapien
758	1	228	0.4	6	P31068	escherichia	831	261	1	YH86_SYNY3	P73335	synecocyst
759	1	228	0.4	6	P50540	mus musculus	832	262	1	PH4H_PSEAE	P43334	pseudomonas
760	1	231	0.4	6	O45551	caenorhabdi	833	262	1	SPRE_RAT	P18297	rattus norv
761	1	231	0.4	6	P26717	homo sapien	834	262	1	TRT3_CHICK	P12620	gallus gall
762	1	233	0.4	6	Q58309	methanococc	835	262	1	Y877_MYCTU	Q10539	mycobacteri
763	1	233	0.4	6	P26715	homo sapien	836	263	1	KDGR_ECOLI	P76268	escherichia



545	1	125	1	H2B_CHICK	P02279	gallus gall	618	6	0.4	162	1	THY1_MOUSE	P01831	mus musculus
546	1	127	1	RL22_XENLA	P50886	xenopus lae	619	6	0.4	163	1	ILVH_HAEIN	P45260	haemophilus
547	1	127	1	VYV1_AZOVI	P40431	azotobacter	620	6	0.4	164	1	ATPX_CVAPA	P48085	cyanophora
548	1	128	1	CHEY_ECOLI	P06143	escherichia	621	6	0.4	164	1	VG59_BPML5	Q05272	mycobacteri
549	1	128	1	CHEY_SALTY	P06657	salmonella	622	6	0.4	165	1	LE24_ARCFU	O28513	archaeoglob
550	1	128	1	YT24_CABEL	Q10933	caenorhabdi	623	6	0.4	165	1	MLCH_HUMAN	P20382	homo sapien
551	1	129	1	A2UR_ALCXX	P56547	alcaligenes	624	6	0.4	165	1	MLCH_RAT	P14200	rattus norv
552	1	129	1	A2UR_ALCSP	P00279	alcaligenes	625	6	0.4	166	1	BIOX_BACSH	P22821	bacillus sp
553	1	129	1	RK12_GUITH	O78414	guillardia	626	6	0.4	166	1	ING_CAPHI	P79154	capra hircu
554	1	129	1	YASD_LEPIN	P41395	leptosira	627	6	0.4	166	1	MLCH_MOUSE	P56942	mus musculus
555	1	130	1	ACR2_YEAST	Q06597	saccharomyc	628	6	0.4	167	1	RNBR_BOVIN	P39873	bos taurus
556	1	130	1	LV1G_HUMAN	P06316	homo sapien	629	6	0.4	168	1	GA15_MOUSE	P35639	mus musculus
557	1	130	1	TAT_HV2BE	P18098	human immu	630	6	0.4	168	1	IA03_WHEAT	P17314	triticum ae
558	1	130	1	TVA1_MOUSE	P01738	mus musculu	631	6	0.4	168	1	TCTP_HVBR	O9ZSW9	hevea bras
559	1	131	1	LYGE_HUMAN	P16553	homo sapien	632	6	0.4	168	1	TCTP_TOBAC	O9XHL7	nicotiana t
560	1	131	1	PSBU_SINY3	Q55332	synecocyst	633	6	0.4	168	1	YE92_HAEIN	P44217	haemophilus
561	1	132	1	FLJS_PSEAE	O33422	pseudomonas	634	6	0.4	169	1	CAFI_DICDI	P54670	dictyostell
562	1	132	1	Y433_METTH	O26533	methanobact	635	6	0.4	171	1	IAAD_HORVU	P11643	hordeum vul
563	1	133	1	TATB_VIBCH	P57063	vibrio chol	636	6	0.4	172	1	RL11_TETTH	P24119	tetrahymena
564	1	133	1	VG4_SPV4	P11336	spiroplasma	637	6	0.4	173	1	FRI3_RANCA	P07797	rana catesb
565	1	134	1	H2AV_YEAST	O12692	saccharomyc	638	6	0.4	174	1	PLU_DROME	P42570	drosophila
566	1	135	1	IL5_CAVPO	Q08987	cavia porce	639	6	0.4	175	1	AXIS_ARATH	Q9Xfmo	arabidopsis
567	1	135	1	RS6_ECOLI	P02358	escherichia	640	6	0.4	176	1	RL10_STRCO	P41103	streptomyce
568	1	137	1	MP12_LYMT	P25289	lymaea sta	641	6	0.4	177	1	GRPE_THETH	Q56236	thermus aqu
569	1	137	1	RET4_HUMAN	P29373	homo sapien	642	6	0.4	177	1	YNIW_AZCH	P23177	azotobacter
570	1	137	1	RK16_ODOSI	P49553	odontella s	643	6	0.4	178	1	LACB_PIG	P04119	sus scrofa
571	1	138	1	MP15_LYMT	P31241	lymaea sta	644	6	0.4	179	1	FLIN_AGRU	Q57259	agrobacteri
572	1	138	1	XJH5_YEAST	P47036	saccharomyc	645	6	0.4	179	1	GRPE_LACLC	P42369	lactococcus
573	1	139	1	R1SB_METTH	O27443	methanobact	646	6	0.4	179	1	SPI7_MONDO	O62771	monodelphis
574	1	140	1	DHSB_ARATH	P21915	arabidopsis	647	6	0.4	179	1	UBC2_CANAL	O74201	candida alb
575	1	140	1	FUCU_ECOLI	P11555	escherichia	648	6	0.4	180	1	GLUC_CAVPO	P05110	cavia porce
576	1	140	1	YP15_STAAU	P13977	staphylococ	649	6	0.4	180	1	YAF5_YEAST	P39718	saccharomyc
577	1	141	1	ALG2_TYRPU	O02380	tyrophagus	650	6	0.4	182	1	HSB2_HUMAN	Q16082	homo sapien
578	1	141	1	YORB_CABEL	Q09338	caenorhabdi	651	6	0.4	182	1	HSB2_RAT	Q35878	rattus norv
579	1	141	1	ISOD_SULAC	P37820	sulfolobus	652	6	0.4	182	1	NODC_BRAEL	P53417	bradyrhizob
580	1	143	1	IF2B_METJA	Q57562	methanococc	653	6	0.4	182	1	Y391_HAEIN	P43992	haemophilus
581	1	144	1	YBGT_PSEAE	O68560	pseudomonas	654	6	0.4	182	1	YGFA_ECOLI	P09160	escherichia
582	1	145	1	HBBI_TRICR	P10785	triturus cr	655	6	0.4	183	1	E321_ADELA	P35769	human adeno
583	1	146	1	OB_CANPA	Q02720	canis famil	656	6	0.4	183	1	VATE_METMA	Q60183	methanosarc
584	1	147	1	MCP_MEDSA	P42067	medicago sa	657	6	0.4	183	1	YGJV_ECOLI	P42603	escherichia
585	1	147	1	SPAM_SALTY	P40612	salmonella	658	6	0.4	184	1	BCM_HUMAN	Q02223	homo sapien
586	1	147	1	UCRH_YEAST	P00127	saccharomyc	659	6	0.4	184	1	KITH_THEMA	O9WYN2	thermotoga
587	1	147	1	Y846_SINY3	Q55433	synecocyst	660	6	0.4	185	1	CBX1_HUMAN	P23197	homo sapien
588	1	147	1	Y180_SINY3	P74125	synecocyst	661	6	0.4	185	1	FLAV_AQUAE	O67866	aquifex aeo
589	1	148	1	BENE_HUMAN	Q13021	homo sapien	662	6	0.4	185	1	PHNN_ECOLI	P16690	escherichia
590	1	148	1	CALM_PARTE	P07463	paramecium	663	6	0.4	185	1	RR5_ODOSI	P49493	odontella s
591	1	148	1	CC21_PEA	P19026	pisum sativ	664	6	0.4	185	1	Y754_METJA	Q58164	methanococc
592	1	148	1	PFDA_PYRHO	O58263	pyrococcus	665	6	0.4	187	1	ATPD_ODOSI	Q00821	odontella s
593	1	148	1	YCBG_HAEIN	P44161	haemophilus	666	6	0.4	187	1	E321_ADELP	P35770	human adeno
594	1	149	1	NDK_SINY3	P74494	synecocyst	667	6	0.4	187	1	E321_ADE35	P15139	human adeno
595	1	149	1	YFYZ_ECOLI	P76538	escherichia	668	6	0.4	189	1	ICYA_MANSE	P00305	manduca sex
596	1	150	1	YP59_YEAST	Q02784	saccharomyc	669	6	0.4	189	1	KADL_METTH	O27699	methanobact
597	1	151	1	YH20_RHOCA	Q02999	rhodobacter	670	6	0.4	189	1	RASK_HUMAN	P01116	homo sapien
598	1	152	1	RECV_VIBCH	Q56647	vibrio chol	671	6	0.4	189	1	RASK_MOUSE	P32883	mus musculu
599	1	152	1	YF47_SINY3	P74662	synecocyst	672	6	0.4	189	1	RASK_MSVKI	P01117	kirsten mur
600	1	152	1	YFJX_ECOLI	P52139	escherichia	673	6	0.4	189	1	YADM_ECOLI	P37018	escherichia
601	1	153	1	Y264_METJA	Q57712	methanococc	674	6	0.4	191	1	CLUS_MSAU	P14683	mesocricetu
602	1	154	1	KMLS_MELGA	P56276	meleagris g	675	6	0.4	191	1	SNGI_MOUSE	Q55100	mus musculus
603	1	154	1	YKOL_CABEL	P34289	caenorhabdi	676	6	0.4	191	1	TF2D_PYRAB	Q9V024	pyrococcus
604	1	155	1	PCP_HAEIN	P10325	haemophilus	677	6	0.4	191	1	TF2D_PYRFO	O57050	pyrococcus
605	1	155	1	YHCH_HAEIN	P44583	haemophilus	678	6	0.4	191	1	TF2D_PYRHO	O58737	pyrococcus
606	1	156	1	RM25_YEAST	P23369	saccharomyc	679	6	0.4	193	1	GNTK_SCHPO	Q10242	schizosacch
607	1	156	1	UBCA_YEAST	P52492	saccharomyc	680	6	0.4	193	1	NUIC_SINY3	P26525	synecocyst
608	1	157	1	UBC3_SCHPO	P40984	schizosacch	681	6	0.4	193	1	Y377_MYCPN	P75223	mycoplasma
609	1	158	1	RHM2_HUMAN	P25791	homo sapien	682	6	0.4	193	1	Y455_METJA	Q57897	methanococc
610	1	158	1	RHM2_MOUSE	P25801	mus musculu	683	6	0.4	193	1	C24A_HUMAN	P13498	homo sapien
611	1	158	1	RL15_AERPE	Q9Yf98	aeropyrum p	684	6	0.4	194	1	PTH_ECOLI	P23932	escherichia
612	1	158	1	YIM2_BPPH1	P10426	bacterioph	685	6	0.4	194	1	PTH_SALT1	Q60001	salmonella
613	1	159	1	PTXA_MYCPN	P75292	mycoplasma	686	6	0.4	194	1	Y041_THEAC	P60001	thermoplasm
614	1	159	1	YHAC_YEAST	P38751	saccharomyc	687	6	0.4	194	1	YD16_HAEIN	P44159	haemophilus
615	1	161	1	THY1_RAT	P01830	rattus norv	688	6	0.4	196	1	NODA_RHIS3	P72329	rhizobium s
616	1	161	1	UREE_PROMI	P17090	proteus mir	689	6	0.4	196	1	NGDH_HAEIN	O57045	haemophilus
617	1	161	1	Y0UD_CABEL	P30631	caenorhabdi	690	6	0.4	196	1	RGS1_HUMAN	Q08116	homo sapien



399	7	0.5	1137	1	MSB1_YEAST	P21339	saccharomyc	472	6	0.4	76	1	ULD1_HCMVA	P16773	human cytom
400	7	0.5	1141	1	CN3A_HUMAN	Q14432	homo sapien	473	6	0.4	77	1	HEMZ_YERPS	Q05338	yersinia ps
401	7	0.5	1141	1	CN3A_RAT	Q62865	rattus norv	474	6	0.4	79	1	Y437_METJA	Q57879	methanococc
402	7	0.5	1152	1	YK00_CAEEL	P34305	caenorhabdi	475	6	0.4	82	1	YFTD_METFE	P65059	methanother
403	7	0.5	1157	1	YXNA_THESA	P36917	thermoanaer	476	6	0.4	83	1	CALD_BOVIN	Q27976	bos taurus
404	7	0.5	1161	1	NRDC_RAT	P47245	rattus norv	477	6	0.4	83	1	CLSS_HAEMA	Q25088	haemophilus ma
405	7	0.5	1181	1	P1B2_HUMAN	Q00722	homo sapien	478	6	0.4	84	1	Y073_NPVOP	O10326	orgyia pseu
406	7	0.5	1214	1	BRF3_HUMAN	Q9ulda	homo sapien	479	6	0.4	87	1	EFTS_ANTSP	Q02855	antithamnio
407	7	0.5	1215	1	SIP1_MOUSE	Q9r0g7	mus musculus	480	6	0.4	87	1	MINE_VIBCH	Q9kqn7	vibrio chol
408	7	0.5	1221	1	PBL2_MOUSE	P37889	mus musculus	481	6	0.4	88	1	YB4D_SCHPO	O13973	schizosacch
409	7	0.5	1227	1	B3A3_MOUSE	P16283	mus musculus	482	6	0.4	88	1	YPB4_ECOLI	P03853	escherichia
410	7	0.5	1227	1	B3A3_RAT	P23348	rattus norv	483	6	0.4	89	1	PE23_SHEEP	Q28550	ovis aries
411	7	0.5	1232	1	B3A3_HUMAN	P48751	homo sapien	484	6	0.4	91	1	YOHU_SERPA	P55757	serratia ma
412	7	0.5	1233	1	B3A3_RABIT	O18917	oryctolagus	485	6	0.4	92	1	DYLI_YEAST	Q02647	saccharomyc
413	7	0.5	1233	1	YF16_YEAST	P43597	saccharomyc	486	6	0.4	92	1	H2B_CRONI	P02280	crocodylus
414	7	0.5	1241	1	TRK1_SACBA	P28569	saccharomyc	487	6	0.4	93	1	LE10_HELAN	P46514	heliathanus
415	7	0.5	1262	1	CVA5_RAT	Q04400	rattus norv	488	6	0.4	93	1	YQIC_CAEEL	Q09283	caenorhabdi
416	7	0.5	1264	1	CVA5_RABIT	P40144	oryctolagus	489	6	0.4	95	1	CXA3_CERLA	P01527	cerebratulu
417	7	0.5	1265	1	CVA5_CANFA	P30803	canis famli	490	6	0.4	96	1	HG14_MOUSE	P18608	mus musculus
418	7	0.5	1361	1	YME9_YEAST	Q04693	saccharomyc	491	6	0.4	96	1	RS18_BORBU	O51140	borrelia bu
419	7	0.5	1377	1	NEO1_RAT	P97603	rattus norv	492	6	0.4	99	1	HMG_TETPY	P40625	tetrahymena
420	7	0.5	1402	1	IF4G_RABIT	P41110	oryctolagus	493	6	0.4	100	1	RS24_METTH	Q26367	methanobact
421	7	0.5	1416	1	BLM_MOUSE	O88700	mus musculus	494	6	0.4	101	1	RS14_CHLPN	Q926w9	chlamydia p
422	7	0.5	1432	1	SK13_YEAST	P17883	saccharomyc	495	6	0.4	101	1	THYP_HUMAN	P20962	homo sapien
423	7	0.5	1432	1	NEO1_CHICK	Q90610	gallus gall	496	6	0.4	103	1	RL21_ECOLI	P04422	escherichia
424	7	0.5	1461	1	NEO1_HUMAN	Q92859	homo sapien	497	6	0.4	103	1	RL21_HAEIN	P44359	haemophilus
425	7	0.5	1493	1	ERC6_HUMAN	Q03468	homo sapien	498	6	0.4	105	1	RL44_YEAST	P02405	saccharomyc
426	7	0.5	1493	1	NEO1_MOUSE	P97798	mus musculus	499	6	0.4	105	1	YJN5_YEAST	P47012	saccharomyc
427	7	0.5	1527	1	CAIH_MOUSE	P39061	mus musculus	500	6	0.4	107	1	YCXQ_PORPU	P51384	porphyra pu
428	7	0.5	1580	1	ACB8_HUMAN	Q09428	homo sapien	501	6	0.4	109	1	LVIF_HUMAN	P04208	homo sapien
429	7	0.5	1581	1	ACC8_CRICR	Q09427	cricketus cr	502	6	0.4	109	1	LVII_HUMAN	P06888	homo sapien
430	7	0.5	1610	1	CCAD_MESAU	Q99244	mesocricetu	503	6	0.4	111	1	LVIIA_HUMAN	P01699	homo sapien
431	7	0.5	1647	1	SN24_HUMAN	P51532	homo sapien	504	6	0.4	111	1	LVIC_HUMAN	P01701	homo sapien
432	7	0.5	1717	1	RPAL_MOUSE	Q05134	mus musculus	505	6	0.4	111	1	LVID_HUMAN	P01702	homo sapien
433	7	0.5	1720	1	FTSH_CHLYU	P56369	chlorella v	506	6	0.4	112	1	YCX1_CHLRE	P05723	chlamydomon
434	7	0.5	1816	1	KF1B_HUMAN	O60333	homo sapien	507	6	0.4	112	1	LVIB_HUMAN	P01700	homo sapien
435	7	0.5	1816	1	KF1B_MOUSE	Q60575	mus musculus	508	6	0.4	113	1	EXP9_STRPN	P35599	streptococc
436	7	0.5	1820	1	CINA_ELEEL	P02719	electrophor	509	6	0.4	115	1	SOR_PYRHO	O58810	pyrococcus
437	7	0.5	1857	1	FAS2_PENPA	P15368	penicillium	510	6	0.4	116	1	RT11_CHOCH	A48941	chondrus cr
438	7	0.5	1902	1	SMFL_HUMAN	O14497	homo sapien	511	6	0.4	116	1	TKNK_MOUSE	P55099	mus musculus
439	7	0.5	1966	1	CCAF_HUMAN	O60840	homo sapien	512	6	0.4	116	1	TKNK_RAT	P08435	rattus norv
440	7	0.5	1982	1	CHDM_DROME	O97159	drosophila	513	6	0.4	116	1	XKDD_BACSU	P39783	bacillus su
441	7	0.5	2009	1	SEC7_YEAST	P11075	saccharomyc	514	6	0.4	117	1	VATG_DROME	O9azh6	drosophila
442	7	0.5	2110	1	MCAS_MYCBO	Q02251	mycobacteri	515	6	0.4	120	1	YEO7_SCHPO	O13804	schizosacch
443	7	0.5	2161	1	CCAD_HUMAN	Q01668	homo sapien	516	6	0.4	120	1	YHIT_RICPR	Q9zd11	ricketsia
444	7	0.5	2203	1	CCAD_RAT	P27732	rattus norv	517	6	0.4	121	1	ANFB_MOUSE	P40753	mus musculus
445	7	0.5	2353	1	CCAH_HUMAN	O95180	homo sapien	518	6	0.4	121	1	RBS_ALVHS	P24682	alvinoconch
446	7	0.5	3135	1	S230_PLAFO	Q08372	plasmodium	519	6	0.4	122	1	H2B1_PSAMI	P02287	psammecchinu
447	7	0.5	3418	1	BRC2_HUMAN	P51587	homo sapien	520	6	0.4	122	1	H2B2_PSAMI	P02288	psammecchinu
448	7	0.5	4385	1	YF73_CAEEL	Q09222	caenorhabdi	521	6	0.4	122	1	H2BL_STRPU	P16888	strongyloce
449	7	0.5	4829	1	BIR6_HUMAN	Q9nr09	homo sapien	522	6	0.4	122	1	H2BN_STRPU	P16889	strongyloce
450	7	0.5	5035	1	RYNR_PIG	P16960	sus scrofa	523	6	0.4	122	1	RL7_NEIMA	P80716	neisseria m
451	7	0.5	5037	1	RYNR_RABIT	P11716	oryctolagus	524	6	0.4	123	1	H2BE_STRPU	P02289	strongyloce
452	6	0.4	12	1	UCRH_MOUSE	P99028	mus musculus	525	6	0.4	123	1	H2B_SALTR	Q02282	salmo trutt
453	6	0.4	37	1	RL36_UREPA	Q9pqn7	ureaplasma	526	6	0.4	124	1	H2B_RAT	Q00715	rattus norv
454	6	0.4	47	1	HMC4_DESVH	P33391	desulfovibr	527	6	0.4	124	1	REV_SIVC2	P17280	chimpanzee
455	6	0.4	63	1	CSRA_HAEIN	P44879	haemophilus	528	6	0.4	125	1	H2B1_MOUSE	P10853	mus musculus
456	6	0.4	63	1	ICE1_ASCSU	P07851	ascaris suu	529	6	0.4	125	1	H2B1_XENLA	P02281	xenopus lae
457	6	0.4	64	1	Y592_HAEIN	P44021	haemophilus	530	6	0.4	125	1	H2B2_MOUSE	P10854	mus musculus
458	6	0.4	69	1	YDHZ_ECOLI	P77274	escherichia	531	6	0.4	125	1	H2B2_XENLA	P06900	xenopus lae
459	6	0.4	70	1	RR18_PORPU	P51256	porphyra pu	532	6	0.4	125	1	H2BA_HUMAN	P02278	homo sapien
460	6	0.4	70	1	VC17_SPVKA	P32218	swinepox vi	533	6	0.4	125	1	H2BC_HUMAN	Q99880	homo sapien
461	6	0.4	70	1	Y650_HAEIN	P44028	haemophilus	534	6	0.4	125	1	H2BD_HUMAN	Q99877	homo sapien
462	6	0.4	71	1	RR18_MESVI	Q9mup6	mesostigma	535	6	0.4	125	1	H2BE_HUMAN	Q99879	homo sapien
463	6	0.4	71	1	YVFE_VACCC	P20563	vaccinia vi	536	6	0.4	125	1	H2BF_HUMAN	P33778	homo sapien
464	6	0.4	72	1	MARB_ECOLI	P31121	escherichia	537	6	0.4	125	1	H2BH_HUMAN	Q93078	homo sapien
465	6	0.4	72	1	RL29_CHLTP	P28538	chlamydia t	538	6	0.4	125	1	H2BJ_HUMAN	Q93079	homo sapien
466	6	0.4	72	1	RR18_ODOSI	P49505	odontella s	539	6	0.4	125	1	H2BL_HUMAN	Q93080	homo sapien
467	6	0.4	73	1	TDCR_ECOLI	P11866	escherichia	540	6	0.4	125	1	H2BN_HUMAN	P23527	homo sapien
468	6	0.4	73	1	PSBH_MARPO	P12160	marchantia	541	6	0.4	125	1	H2BQ_HUMAN	Q16778	homo sapien
469	6	0.4	75	1	MT1_WHEAT	P43400	tritium ae	542	6	0.4	125	1	H2BR_HUMAN	P06899	homo sapien
470	6	0.4	75	1	PSAE_SYNEN	P25898	synechococc	543	6	0.4	125	1	H2BS_HUMAN	P57053	homo sapien
471	6	0.4	75	1	RS18_RHOCA	O68127	rhodobacter	544	6	0.4	125	1	H2B_CAIMO	P14001	cairina mos



253	1	GL9P_HUMAN	1	527	7	0.5	326	706	1	NUCL_HUMAN	P19338	homo sapien
254	1	ACH2_CHICK	1	528	7	0.5	327	706	1	NUCL_MOUSE	P09405	mus musculus
255	1	CALD_RAT	1	531	7	0.5	328	708	1	ABBI_MOUSE	O9qx11	mus musculus
256	1	C7C4_MAIZE	1	538	7	0.5	329	713	1	NUCL_MESAU	P08199	mesocricetu
257	1	TULLI_HUMAN	1	542	7	0.5	330	716	1	BAC2_MOUSE	P97303	mus musculus
258	1	PAK3_HUMAN	1	544	7	0.5	331	721	1	YCF2_OENPI	P31568	oenothera p
259	1	PAK3_MOUSE	1	544	7	0.5	332	728	1	CTCF_CHICK	Q08705	gallus gall
260	1	PAK3_RAT	1	544	7	0.5	333	752	1	DRS1_YEAST	P32892	saccharomyc
261	1	THS_METKA	1	545	7	0.5	334	759	1	SCT1_YEAST	P32784	saccharomyc
262	1	IF37_MOUSE	1	547	7	0.5	335	763	1	APP2_HUMAN	Q06481	homo sapien
263	1	IF37_HUMAN	1	548	7	0.5	336	763	1	YN51_YEAST	P42843	saccharomyc
264	1	YBOO_YEAST	1	552	7	0.5	337	765	1	APP2_RAT	P15943	rattus norv
265	1	PEX2_PODAN	1	554	7	0.5	338	765	1	UBF1_MOUSE	P25976	mus musculus
266	1	ECR_MANSE	1	556	7	0.5	339	777	1	RGL2_HUMAN	O15211	homo sapien
267	1	INX7_CAEEL	1	556	7	0.5	340	780	1	VAC1_HUMAN	Q93034	homo sapien
268	1	PP71_HCMVA	1	559	7	0.5	341	780	1	VAC1_RABIT	Q29425	oryctolagus
269	1	SNTC_HUMAN	1	561	7	0.5	342	780	1	VAC1_RAT	O9fj31	rattus norv
270	1	MXID-SHIFL	1	566	7	0.5	343	786	1	PRGR_CHICK	P07812	gallus gall
271	1	MXID-SHISO	1	566	7	0.5	344	792	1	UBPA_YEAST	P53874	saccharomyc
272	1	NCAP_MOPEI	1	570	7	0.5	345	795	1	Y210_HUMAN	O92609	homo sapien
273	1	TERM_ADEG1	1	575	7	0.5	346	799	1	CN3B_MOUSE	O61409	mus musculus
274	1	DPO4_YEAST	1	581	7	0.5	347	802	1	NAB3_YEAST	P38996	saccharomyc
275	1	SYN2_RAT	1	586	7	0.5	348	809	1	LEF_BACAN	P15917	bacillus an
276	1	TRM1_ARATH	1	586	7	0.5	349	809	1	UBP1_YEAST	P25037	saccharomyc
277	1	GOX_PENAG	1	587	7	0.5	350	817	1	PPSA_PYRFO	P42850	pyrococcus
278	1	RGPI_HUMAN	1	587	7	0.5	351	819	1	PPSA_PYRAB	O9v2h7	pyrococcus
279	1	YM72_YEAST	1	590	7	0.5	352	819	1	SWEI_YEAST	P32944	saccharomyc
280	1	YN48_YEAST	1	591	7	0.5	353	821	1	PPSA_PYRHO	O57830	pyrococcus
281	1	LAM2_MOUSE	1	592	7	0.5	354	824	1	ROU_HUMAN	Q00839	homo sapien
282	1	CENB_HUMAN	1	599	7	0.5	355	825	1	RCAL_YEAST	P40341	saccharomyc
283	1	CENB_MOUSE	1	599	7	0.5	356	852	1	SRCH_RABIT	P16230	oryctolagus
284	1	DRIL_MOUSE	1	601	7	0.5	357	867	1	PROM_MOUSE	O54990	mus musculus
285	1	PEHX_ERWCH	1	602	7	0.5	358	872	1	SL45_HUMAN	Q13435	homo sapien
286	1	GOX_TALFL	1	605	7	0.5	359	886	1	CN4A_HUMAN	P27815	homo sapien
287	1	CAC1_YEAST	1	606	7	0.5	360	886	1	LEUR_YEAST	P08638	saccharomyc
288	1	CENB_CRIGR	1	606	7	0.5	361	886	1	NIAL_SOYBN	P54233	glycine max
289	1	KU70_MOUSE	1	608	7	0.5	362	886	1	SUHW_DROAN	Q08875	drosophila
290	1	ABPI_SACEX	1	617	7	0.5	363	890	1	NIA2_SOYBN	P39870	glycine max
291	1	SPPA_ECOLI	1	618	7	0.5	364	891	1	MAZ3_SCHCO	P37937	schizophyll
292	1	YKRA_YEAST	1	618	7	0.5	365	899	1	SUHW_DROVI	Q08876	drosophila
293	1	GG95_HUMAN	1	620	7	0.5	366	909	1	CNG4_HUMAN	Q14028	homo sapien
294	1	YD4B_SCHPO	1	629	7	0.5	367	912	1	PGCB_BOVIN	Q28062	bos taurus
295	1	YCF2_OENVI	1	630	7	0.5	368	919	1	PWP2_HUMAN	O13269	homo sapien
296	1	SL56_HUMAN	1	635	7	0.5	369	924	1	KEX2_CANAL	O13359	candida alb
297	1	NTPR_HUMAN	1	636	7	0.5	370	929	1	RBMA_HUMAN	P98175	homo sapien
298	1	NFM_RABIT	1	644	7	0.5	371	932	1	YATA_SCHPO	Q09897	schizosacch
299	1	PDA4_HUMAN	1	645	7	0.5	372	960	1	FGDI_MOUSE	P52734	mus musculus
300	1	WEEL_HUMAN	1	646	7	0.5	373	960	1	VP41_LYCES	P93231	lycopersico
301	1	WEEL_MOUSE	1	646	7	0.5	374	960	1	Y682_HUMAN	O9y4c8	homo sapien
302	1	DREB_CHICK	1	652	7	0.5	375	963	1	REF1_MOUSE	P48377	mus musculus
303	1	C1K4_HUMAN	1	653	7	0.5	376	967	1	HEX_AD802	P03277	human adeno
304	1	C1K4_MOUSE	1	654	7	0.5	377	971	1	NAM7_YEAST	P30771	saccharomyc
305	1	C1K4_MOUSEPF	1	654	7	0.5	378	975	1	KINH_DROME	P17210	drosophila
306	1	C1K4_RAT	1	655	7	0.5	379	975	1	VP41_ARATH	P93043	arabidopsis
307	1	NTPR_RAT	1	661	7	0.5	380	979	1	REF1_HUMAN	P22670	homo sapien
308	1	TERM_ADEB3	1	663	7	0.5	381	999	1	OXRP_CRIGR	O60432	cricetus
309	1	FILS_HUMAN	1	665	7	0.5	382	999	1	OXRE_RAT	O63617	rattus norv
310	1	SSRP_CHICK	1	669	7	0.5	383	1034	1	CAPP_SYNY3	P74299	synecocyst
311	1	NUOL_NEIMA	1	674	7	0.5	384	1041	1	DD16_HUMAN	O60231	homo sapien
312	1	NUOL_NEIMB	1	674	7	0.5	385	1043	1	TCF8_MESAU	O60542	mesocricetu
313	1	VPS5_YEAST	1	675	7	0.5	386	1058	1	PMAL_DICDI	P54679	dictyosteli
314	1	C1CL_RABIT	1	678	7	0.5	387	1062	1	SUM1_YEAST	P46676	saccharomyc
315	1	GARP_PLAFF	1	678	7	0.5	388	1067	1	SGG_DROME	P18431	drosophila
316	1	YHC9_YEAST	1	679	7	0.5	389	1078	1	CYA7_BOVIN	Q29450	bos taurus
317	1	RPSD_RHIME	1	684	7	0.5	390	1080	1	CYA7_HUMAN	P51828	homo sapien
318	1	CICK_HUMAN	1	687	7	0.5	391	1099	1	CYA7_MOUSE	P51829	mus musculus
319	1	CICK_RABIT	1	687	7	0.5	392	1108	1	CN3B_RAT	O63085	rattus norv
320	1	CICK_RAT	1	687	7	0.5	393	1109	1	TCF8_RAT	O62947	rattus norv
321	1	C1CL_RAT	1	687	7	0.5	394	1113	1	HDA5_MOUSE	O92246	mus musculus
322	1	YK06_YEAST	1	692	7	0.5	395	1114	1	TCF8_CHICK	P36197	gallus gall
323	1	APP2_MOUSE	1	695	7	0.5	396	1117	1	TCF8_MOUSE	O64318	mus musculus
324	1	SRCH_HUMAN	1	699	7	0.5	397	1123	1	RBMB_HUMAN	P78332	homo sapien
325	1	HPS_MOUSE	1	704	7	0.5	398	1124	1	TCF8_HUMAN	P37275	homo sapien



107	7	0.5	267	1	XPA_CHICK	P27089	gallus gall	180	7	0.5	406	1	GCST_SOLTU	P54260	solanum tu
108	7	0.5	268	1	RS3A_ANOGA	P52813	anopheles g	181	7	0.5	406	1	MYC_BRARE	P52160	brachydanic
109	7	0.5	272	1	CGHL_MOUSE	P53702	mus musculus	182	7	0.5	411	1	FKB3_YEAST	P38911	saccharomyc
110	7	0.5	273	1	XPA_HUMAN	P20205	homo sapien	183	7	0.5	411	1	MP62_LYTP1	P51753	lytechinus
111	7	0.5	278	1	TRT3_RAB1T	P02641	oryctolagus	184	7	0.5	412	1	FBM4_HUMAN	P57775	homo sapien
112	7	0.5	278	1	YD87_YEAST	P35178	saccharomyc	185	7	0.5	412	1	FKB4_SPOFR	Q26486	spodoptera
113	7	0.5	279	1	ASF1_YEAST	P32447	saccharomyc	186	7	0.5	416	1	RAGE_BOVIN	Q28173	bos taurus
114	7	0.5	280	1	Y326_THEMA	Q9wyg1	thermotoga	187	7	0.5	423	1	HRP_PLABE	Q08168	plasmodium
115	7	0.5	282	1	YKGG_ECOLI	P77433	escherichia	188	7	0.5	427	1	YKR2_YEAST	P36049	saccharomyc
116	7	0.5	288	1	SSRA_ONCMY	P45433	oncorhynchu	189	7	0.5	430	1	GNL1_HUMAN	P36915	homo sapien
117	7	0.5	291	1	CAH8_MOUSE	P28651	mus musculus	190	7	0.5	430	1	GNL1_MOUSE	P36916	mus musculus
118	7	0.5	292	1	Y152_HUMAN	Q14165	homo sapien	191	7	0.5	433	1	KSGC_RAT	P55205	rattus norv
119	7	0.5	293	1	RR4_SPIO	Q49937	spinacia ol	192	7	0.5	435	1	CGS5_YEAST	P30283	saccharomyc
120	7	0.5	295	1	CGD1_HUMAN	P24385	homo sapien	193	7	0.5	439	1	LIP1_DROME	Q98937	drosophila
121	7	0.5	295	1	YDT5_SCHPO	O14210	schizosacch	194	7	0.5	440	1	FXGA_CHICK	O98937	gallus gall
122	7	0.5	299	1	YPX_CRIGR	Q60415	cricetulus	195	7	0.5	445	1	TBB4_WHEAT	Q92ra9	trititum ae
123	7	0.5	299	1	YPX_HUMAN	Q60855	homo sapien	196	7	0.5	445	1	TBB7_MAIZE	Q41784	zea mays (m
124	7	0.5	300	1	STAL_PIG	P31936	sus scrofa	197	7	0.5	446	1	CMGA_PIG	P04404	sus scrofa
125	7	0.5	302	1	BIRA_HAEIN	P46363	haemophilus	198	7	0.5	446	1	YDUG_HUMAN	Q95544	homo sapien
126	7	0.5	303	1	YPY3_SCHPO	P32587	schizosacch	199	7	0.5	448	1	A2AB_CAVPO	Q60475	cavia porce
127	7	0.5	304	1	CEC1_CAEEL	P34618	caenorhabdi	200	7	0.5	448	1	AAC2_DICD1	PL4196	dictyosteli
128	7	0.5	309	1	KDGG_ECOLI	P37647	escherichia	201	7	0.5	449	1	ME53_NPVAC	Q04719	autographa
129	7	0.5	310	1	STAL_BOVIN	Q28862	bos taurus	202	7	0.5	449	1	TBB2_SOYBN	P12460	glycine max
130	7	0.5	312	1	YCE9_YEAST	P25577	saccharomyc	203	7	0.5	450	1	A2AB_HUMAN	P18089	homo sapien
131	7	0.5	315	1	5X22_HUMAN	O15370	homo sapien	204	7	0.5	452	1	AAMP_HUMAN	Q13685	homo sapien
132	7	0.5	316	1	RLA0_PLAF8	Q94660	plasmodium	205	7	0.5	452	1	P185_HUMAN	O15231	homo sapien
133	7	0.5	317	1	RNHL_NEUCR	Q9p5x8	neurospora	206	7	0.5	454	1	P2CH_HUMAN	P49593	homo sapien
134	7	0.5	317	1	STAL1_HUMAN	P21815	homo sapien	207	7	0.5	455	1	HYIN_PSESS	P06618	pseudomonas
135	7	0.5	320	1	HLPL_HUMAN	P54257	homo sapien	208	7	0.5	457	1	CMGA_HUMAN	P10645	homo sapien
136	7	0.5	320	1	STAL_RAT	P13839	rattus norv	209	7	0.5	460	1	NPL2_MOUSE	P1860	mus musculus
137	7	0.5	324	1	STAL_MOUSE	Q61711	mus musculus	210	7	0.5	462	1	BIND_LYTVA	P23117	lytechinus
138	7	0.5	328	1	RPOA_SHEPS	P74963	shewanella	211	7	0.5	463	1	RFBB_VIBCH	Q06951	vibrio chol
139	7	0.5	329	1	LEG6_VICFA	P16079	vicia faba	212	7	0.5	464	1	LEO1_YEAST	P38439	saccharomyc
140	7	0.5	329	1	RPOA_BUCAI	P57566	buchnera ap	213	7	0.5	467	1	SE54_YEAST	Q02825	saccharomyc
141	7	0.5	329	1	RPOA_ECOLI	P00574	escherichia	214	7	0.5	467	1	SYE_BUCAI	P57173	buchnera ap
142	7	0.5	329	1	RPOA_SHEVI	Q950q8	shewanella	215	7	0.5	470	1	BFRI_YEAST	P38934	saccharomyc
143	7	0.5	329	1	Y526_CHLPH	Q92826	chlamydia p	216	7	0.5	471	1	FBX3_HUMAN	Q9uk99	homo sapien
144	7	0.5	330	1	NDFM_CHICK	P79766	gallus gall	217	7	0.5	471	1	PM22_DROVI	Q24705	drosophila
145	7	0.5	331	1	MAZ_MESAU	P56670	mesocricetu	218	7	0.5	474	1	KCC4_RAT	P13234	rattus norv
146	7	0.5	333	1	RPOA_PSEAE	O52760	pseudomonas	219	7	0.5	474	1	LAM3_MOUSE	P48680	mus musculus
147	7	0.5	335	1	LEG2_VICFA	P16078	vicia faba	220	7	0.5	475	1	YMF4_MOUSE	O04213	saccharomyc
148	7	0.5	336	1	HK23_MOUSE	P97334	mus musculus	221	7	0.5	477	1	MAZ_HUMAN	P56270	homo sapien
149	7	0.5	336	1	RLA0_METHF	O27717	methanobact	222	7	0.5	477	1	MAZ_MOUSE	P56671	mus musculus
150	7	0.5	336	1	RUVB_HELPY	Q92m57	helicobacte	223	7	0.5	478	1	YJM3_YEAST	P47018	saccharomyc
151	7	0.5	336	1	RUVB_HELPY	O25699	helicobacte	224	7	0.5	480	1	BIND_ARBP	P24608	arabacia pun
152	7	0.5	338	1	LEGB_PEA	P14594	pisum sativ	225	7	0.5	481	1	BIND_STRPU	P06651	strongyloce
153	7	0.5	340	1	YMTA_CAEEL	P34486	caenorhabdi	226	7	0.5	482	1	ODB2_BOVIN	P11181	bos taurus
154	7	0.5	344	1	FSA_MOUSE	P47931	mus musculus	227	7	0.5	482	1	ODB2_HUMAN	P11182	homo sapien
155	7	0.5	344	1	FSA_RAT	P21674	rattus norv	228	7	0.5	482	1	ODB2_MOUSE	P53395	mus musculus
156	7	0.5	350	1	LEGG_PEA	P05693	pisum sativ	229	7	0.5	493	1	CPEI_RABIT	P36100	saccharomyc
157	7	0.5	356	1	NDF1_HUMAN	Q13562	homo sapien	230	7	0.5	484	1	LEG4_VICFA	P05190	vicia faba
158	7	0.5	357	1	NDF1_CHICK	P79765	gallus gall	231	7	0.5	485	1	BIND_STRFN	P23118	strongyloce
159	7	0.5	357	1	NDF1_MOUSE	Q60867	mus musculus	232	7	0.5	485	1	SR54_SYNP7	Q55311	synchococc
160	7	0.5	357	1	NDF1_RAT	Q64289	rattus norv	233	7	0.5	491	1	YIJO_YEAST	P40499	saccharomyc
161	7	0.5	368	1	IDH2_KLJLA	O94230	kluyveromyc	234	7	0.5	493	1	GLCI_RABIT	P08682	oryctolagus
162	7	0.5	368	1	MYCL_MOUSE	P10166	mus musculus	235	7	0.5	495	1	GLCI_SOYBN	P04776	glycine max
163	7	0.5	371	1	DEK1_CAEEL	O19992	caenorhabdi	236	7	0.5	497	1	IRF5_MOUSE	P56477	mus musculus
164	7	0.5	375	1	DEK_HUMAN	P35659	homo sapien	237	7	0.5	502	1	ATPA_BACME	P17674	bacillus me
165	7	0.5	375	1	Y385_METTA	O57830	methanococc	238	7	0.5	502	1	ATPA_BACP3	P09219	bacillus ps
166	7	0.5	382	1	AZAB_DIDMA	Q77715	dideiphis m	239	7	0.5	502	1	ATPA_BACST	P42005	bacillus st
167	7	0.5	382	1	NDF2_HUMAN	Q15784	homo sapien	240	7	0.5	503	1	LEG3_PEA	P05692	pisum sativ
168	7	0.5	382	1	NDF2_RAT	Q63689	rattus norv	241	7	0.5	504	1	AINX_MOUSE	P46660	mus musculus
169	7	0.5	382	1	SX11_XENLA	Q91731	xenopus lae	242	7	0.5	504	1	CP42_RAT	P20816	rattus norv
170	7	0.5	383	1	DEMA_HUMAN	Q08495	homo sapien	243	7	0.5	504	1	IRF5_HUMAN	Q13568	homo sapien
171	7	0.5	383	1	NDF2_MOUSE	Q62414	mus musculus	244	7	0.5	505	1	AINX_RAT	P23565	rattus norv
172	7	0.5	392	1	FKB4_YEAST	Q06205	saccharomyc	245	7	0.5	507	1	CP43_RAT	P20817	rattus norv
173	7	0.5	394	1	AZAB_RABIT	O77830	oryctolagus	246	7	0.5	511	1	ACH2_RAT	P12389	rattus norv
174	7	0.5	394	1	HGHI_YEAST	P48362	saccharomyc	247	7	0.5	515	1	C6D4_DROME	Q9vcw1	drosophila
175	7	0.5	394	1	MYCL_CYPCA	Q90341	cypripus ca	248	7	0.5	516	1	ZG53_XENLA	P18728	xenopus lae
176	7	0.5	396	1	TFT_DROME	P19351	drosophila	249	7	0.5	517	1	PLIN_RAT	P43884	rattus norv
177	7	0.5	399	1	MYC_CARAU	P49709	carassius a	250	7	0.5	518	1	IRVA_CAEEL	Q27519	caenorhabdi
178	7	0.5	400	1	ENPL_MESAU	P08712	mesocricetu	251	7	0.5	521	1	IMAI_XENLA	P52170	xenopus lae
179	7	0.5	401	1	MYC2_CYPCA	Q90342	cypripus ca	252	7	0.5	522	1	IMA2_XENLA	P52171	xenopus lae



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:05:29 ; Search time 16.55 seconds  
(without alignments)  
2879.117 Million cell updates/sec

Title: US-09-512-581-2

Perfect score: 1391

Sequence: 1 MAHSTRINDCKITPPGVK.....QKGRGSPSKTPSPQPKKNV 1391

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.6	439	1 TRF1_HUMAN	P54274 homo sapien
2	9	0.6	1938	1 MYSA_MOUSE	Q02566 mus musculus
3	9	0.6	1938	1 MYSA_MOUSE	Q02563 rattus norv
4	8	0.6	87	1 GBG_LOLFO	Q01821 loligo forb
5	8	0.6	91	1 UCRH_HUMAN	P07919 homo sapien
6	8	0.6	181	1 VE4_HPV04	Q07852 human papill
7	8	0.6	206	1 HMG2_CHICK	P26584 gallus gall
8	8	0.6	233	1 EMB1_CAVPO	P22032 cavia porce
9	8	0.6	233	1 SOML_ONCKE	P24405 oncorhynch
10	8	0.6	242	1 THYL_HUMAN	P20396 homo sapien
11	8	0.6	316	1 YNFB_YEAST	P53947 saccharomyc
12	8	0.6	339	1 RLAO_ARCFU	Q28781 archaeoglob
13	8	0.6	418	1 YSE2_CABEL	Q09936 caenorhabdi
14	8	0.6	451	1 GP1D_CHLTR	P10555 chlamydia t
15	8	0.6	463	1 CMGA_MOUSE	P26339 mus musculus
16	8	0.6	466	1 CMGA_MOUSE	P10354 rattus norv
17	8	0.6	513	1 NIEK_RHLSN	P19067 rhizobium s
18	8	0.6	533	1 LCP2_HUMAN	Q13094 homo sapien
19	8	0.6	544	1 INVO_AOTTR	P24708 aotus trivi
20	8	0.6	544	1 NFL_XENLA	P35616 xenopus lae
21	8	0.6	585	1 LAMI_HUMAN	P20700 homo sapien
22	8	0.6	587	1 LAMI_MOUSE	P14733 mus musculus
23	8	0.6	653	1 APPL_MOUSE	Q03157 mus musculus
24	8	0.6	823	1 UBPG_HUMAN	Q95455 homo sapien
25	8	0.6	1002	1 IF2P_YEAST	P39730 saccharomyc
26	8	0.6	1081	1 SPS2_CRAPL	Q04933 craterostig
27	8	0.6	1165	1 CYA6_CANFA	P30804 canis fami
28	8	0.6	1165	1 CYA6_MOUSE	Q01341 mus musculus
29	8	0.6	1166	1 CYA6_RAT	Q03343 rattus norv
30	8	0.6	1168	1 CYA6_HUMAN	Q43306 homo sapien
31	8	0.6	1184	1 FBL2_HUMAN	P98095 homo sapien
32	8	0.6	1220	1 IF2P_HUMAN	Q60841 homo sapien
33	8	0.6	1859	1 GBF1_HUMAN	Q92538 homo sapien

1	MYSB_MESAU	1	1934	0.6	8	34	PL3540 mesocricetu
1	MYSB_RAT	1	1935	0.6	8	35	P02564 rattus norv
1	CYAL_DROME	1	2248	0.6	8	36	P32870 drosophila
1	ATRX_HUMAN	1	2375	0.6	8	37	P46100 homo sapien
1	RR18_GUITH	73	73	0.5	7	38	O78488 guillardia
1	UCRH_BOVIN	78	78	0.5	7	39	P00126 bos taurus
1	E111_ADEM1	96	96	0.5	7	40	P12533 mouse adeno
1	VE7_COPV	97	97	0.5	7	41	O89759 canine oral
1	SNCS_HUMAN	98	98	0.5	7	42	O75971 homo sapien
1	RLI2_METTH	101	101	0.5	7	43	P05394 methanobact
1	THYP_BOVIN	101	101	0.5	7	44	P08814 bos taurus
1	THYP_RAT	101	101	0.5	7	45	P04550 rattus norv
1	HSP3_MOUSE	102	102	0.5	7	46	O62100 mus musculu
1	HXA7_RAT	105	105	0.5	7	47	P09634 rattus norv
1	RLI2_ARCFU	106	106	0.5	7	48	O28780 archaeoglob
1	THYA_BOVIN	109	109	0.5	7	49	P01252 bos taurus
1	THYA_HUMAN	110	110	0.5	7	50	P06454 homo sapien
1	THYA_MOUSE	110	110	0.5	7	51	P26350 mus musculu
1	THVA_RAT	111	111	0.5	7	52	P06302 rattus norv
1	RLA2_PLAFA	112	112	0.5	7	53	O00806 plasmodium
1	GCAD_BACCL	116	116	0.5	7	54	P42817 bacillus ca
1	SMS2_CARAU	120	120	0.5	7	55	O9ygn4 carassius a
1	RBFA_BORBU	127	127	0.5	7	56	O51742 borrelia bu
1	RS19_ARCFU	133	133	0.5	7	57	O28358 archaeoglob
1	CLX2_HUMAN	134	134	0.5	7	58	Q13329 homo sapien
1	IPPD_PIG	137	137	0.5	7	59	O29277 sus scrofa
1	YN8E_YEAST	142	142	0.5	7	60	P53737 saccharomyc
1	HMG2_SOYBN	152	152	0.5	7	61	Q10370 glycine max
1	PURR_RHOCA	154	154	0.5	7	62	O52710 rhodobacter
1	E1BS_ADE40	167	167	0.5	7	63	P10543 human adeno
1	E1BS_ADE41	170	170	0.5	7	64	P10544 human adeno
1	SODC_ECOLI	173	173	0.5	7	65	P53635 escherichia
1	HMG2_SOYBN	176	176	0.5	7	66	Q00423 glycine max
1	HMG1_CRIGR	180	180	0.5	7	67	P07136 cricetus
1	E320_ADE1A	181	181	0.5	7	68	P35767 human adeno
1	E320_ADE1P	181	181	0.5	7	69	P35768 human adeno
1	E320_ADE35	181	181	0.5	7	70	P15137 human adeno
1	Y314_ARATH	182	182	0.5	7	71	O9m83 arabidopsis
1	UBC4_WHEAT	184	184	0.5	7	72	P16577 triticum ae
1	NUPL_XENLA	200	200	0.5	7	73	P05221 xenopus lae
1	IPPD_BOVIN	202	202	0.5	7	74	P07516 bos taurus
1	IPPD_HUMAN	204	204	0.5	7	75	O9ud71 homo sapien
1	6B_AGRVI	206	206	0.5	7	76	O04551 agrobacteri
1	RHTC_ECOLI	206	206	0.5	7	77	P27846 escherichia
1	S3AE_BACSU	206	206	0.5	7	78	P49783 bacillus su
1	HMG2_HUMAN	208	208	0.5	7	79	P26583 homo sapien
1	HMG2_MOUSE	209	209	0.5	7	80	P30681 mus musculu
1	HMG2_PIG	209	209	0.5	7	81	P17741 sus scrofa
1	HMG2_RAT	209	209	0.5	7	82	P52925 rattus norv
1	VDGL_BACSU	209	209	0.5	7	83	P96707 bacillus su
1	RGS2_HUMAN	211	211	0.5	7	84	P41220 homo sapien
1	NEUM_CARAU	213	213	0.5	7	85	P17691 carassius a
1	VE4_HPV65	213	213	0.5	7	86	Q07873 human papil
1	HMG1_MOUSE	214	214	0.5	7	87	P07155 mus musculu
1	CALL_ARATH	215	215	0.5	7	88	P30188 arabidopsis
1	HB7B_XENLA	220	220	0.5	7	89	P04476 xenopus lae
1	TUR8_MOUSE	224	224	0.5	7	90	P19473 mus musculu
1	HT14_ARATH	225	225	0.5	7	91	P46685 arabidopsis
1	EMBP_RAT	227	227	0.5	7	92	O63189 rattus norv
1	HXA7_MOUSE	229	229	0.5	7	93	P02830 mus musculu
1	FSA_XENLA	233	233	0.5	7	94	P31515 xenopus lae
1	EMB2_CAVPO	234	234	0.5	7	95	P35709 cavia porce
1	VGLG_HSV11	238	238	0.5	7	96	P06484 herpes simp
1	CENB_SHEEP	239	239	0.5	7	97	P49451 ovis aries
1	TRPD_BACCA	242	242	0.5	7	98	P30525 bacillus ca
1	EFTS_SPIPL	247	247	0.5	7	99	P34828 spirulina p
1	MTRA_METKA	248	248	0.5	7	100	O32867 methanopyru
1	UL07_HSV6Z	249	249	0.5	7	101	P52457 human herpe
1	YKJ9_YEAST	256	256	0.5	7	102	P34247 saccharomyc
1	RS3A_ORISA	261	261	0.5	7	103	P49397 oryza sativ
1	VP31_FRG3V	261	261	0.5	7	104	P18178 frog virus
1	TRPA_RHOSH	263	263	0.5	7	105	O9x468 rhodobacter
1	EN16_STRPU	265	265	0.5	7	106	PL3665 strongyloce



PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 307;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 74 eeeeeer 81

Search completed: September 25, 2001, 10:06:18  
Job time: 184 sec



RESULT 50  
AAG37417  
ID AAG37417 standard; Protein; 307 AA.  
XX  
AC AAG37417;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46004.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.



PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157533.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 307;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1325 EEEEEEE 1332  
DB 74 eeeeeer 81



05-AUG-1999; 99US-0147192.  
05-AUG-1999; 99US-0147260.  
06-AUG-1999; 99US-0147303.  
06-AUG-1999; 99US-0147416.  
09-AUG-1999; 99US-0147493.  
09-AUG-1999; 99US-0147935.  
10-AUG-1999; 99US-0148171.  
11-AUG-1999; 99US-0148319.  
12-AUG-1999; 99US-0148341.  
12-AUG-1999; 99US-0148565.  
13-AUG-1999; 99US-0148684.  
16-AUG-1999; 99US-0149368.  
17-AUG-1999; 99US-0149175.  
18-AUG-1999; 99US-0149426.  
20-AUG-1999; 99US-0149722.  
20-AUG-1999; 99US-0149723.  
20-AUG-1999; 99US-0149929.  
23-AUG-1999; 99US-0149902.  
23-AUG-1999; 99US-0149930.  
25-AUG-1999; 99US-0150566.  
26-AUG-1999; 99US-0150884.  
27-AUG-1999; 99US-0151065.  
27-AUG-1999; 99US-0151066.  
27-AUG-1999; 99US-0151080.  
30-AUG-1999; 99US-0151303.  
31-AUG-1999; 99US-0151438.  
01-SEP-1999; 99US-0151930.  
07-SEP-1999; 99US-0152363.  
10-SEP-1999; 99US-0153070.  
13-SEP-1999; 99US-0153758.  
15-SEP-1999; 99US-0154018.  
16-SEP-1999; 99US-0154039.  
20-SEP-1999; 99US-0154779.  
22-SEP-1999; 99US-0155139.  
23-SEP-1999; 99US-0155486.  
24-SEP-1999; 99US-0155659.  
28-SEP-1999; 99US-0156458.  
29-SEP-1999; 99US-0156596.  
04-OCT-1999; 99US-0157117.  
05-OCT-1999; 99US-0157753.  
06-OCT-1999; 99US-0157865.  
07-OCT-1999; 99US-0158029.  
08-OCT-1999; 99US-0158232.  
12-OCT-1999; 99US-0158369.  
13-OCT-1999; 99US-0159293.  
13-OCT-1999; 99US-0159294.  
13-OCT-1999; 99US-0159295.  
14-OCT-1999; 99US-0159329.  
14-OCT-1999; 99US-0159330.  
14-OCT-1999; 99US-0159331.  
14-OCT-1999; 99US-0159637.  
14-OCT-1999; 99US-0159638.  
18-OCT-1999; 99US-0159584.  
21-OCT-1999; 99US-0160741.  
21-OCT-1999; 99US-0160767.  
21-OCT-1999; 99US-0160768.  
21-OCT-1999; 99US-0160770.  
21-OCT-1999; 99US-0160814.  
21-OCT-1999; 99US-0160815.  
22-OCT-1999; 99US-0160980.  
22-OCT-1999; 99US-0160981.  
22-OCT-1999; 99US-0160989.  
25-OCT-1999; 99US-0161404.  
25-OCT-1999; 99US-0161405.  
25-OCT-1999; 99US-0161406.  
26-OCT-1999; 99US-0161359.  
26-OCT-1999; 99US-0161360.  
26-OCT-1999; 99US-0161361.  
28-OCT-1999; 99US-0161920.  
28-OCT-1999; 99US-0161992.  
28-OCT-1999; 99US-0161993.  
29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 294;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1324 PEEEEEE 1331  
Db 226 peeeeeee 233  
|||||||  
RESULT 49  
AAG16481  
ID AAG16481 standard; Protein; 307 AA.  
XX  
AC AAG16481;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17144.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135359.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.



PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 260;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

OY 1325 EEEEEER 1332  
Db 27 eeeeeer 34

## RESULT 48

AAAG31669  
ID AAG31669 standard; Protein; 294 AA.

XX AC AAG31669;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38071.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.



PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.



```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 260;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1325 EEEEEER 1332
Db 27 eeeeeer 34

RESULT 47
AAG38888
ID AAG38888 standard; Protein; 260 AA.
XX
AC AAG38888;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48041.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
11-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
```



PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 23-OCT-1999; 99US-0161404.  
 PR 23-OCT-1999; 99US-0161405.  
 PR 23-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162143.  
 PR 29-OCT-1999; 99US-0162144.

Query Match 0.6%; Score 8; DB 21; Length 254;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 21 eeeeeer 28

RESULT 46

AAG08527

ID AAG08527 standard; Protein; 260 AA.

XX AC AAG08527;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6099.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX FN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 23-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144325.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.



PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.



PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144634.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 254;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 1325 EEEEEER 1332  
|||||||  
Db 21 eeeeeer 28

## RESULT 45

AAG37418  
ID AAG37418 standard; Protein; 254 AA.

AC AAG37418;

XX 18-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 46005.

DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.



DR WPI; 2001-182940/18.  
 DR N-PSDB; AAF30409.  
 XX  
 XX New isolated or recombinant polynucleotide encoding an antigenic  
 PT polypeptide (cytokine IL-D80) for providing reagents to regulate  
 PT activation, development, differentiation and function of cells  
 PT including haematopoietic cells -  
 XX  
 XX Claim 1; Page 9; 73pp; English.  
 PS  
 XX The present sequence is that of a novel human cytokine, termed  
 CC variant interleukin DNAX 80 (IL-D80), which shows homology  
 CC to interleukin-11. The sequence was deduced from isolated IL-D80  
 CC cDNA (see AAF30409). A human IL-D80 sequence is also provided in  
 CC AAB20275. The 2 IL-D80 sequences differ only in the signal peptide  
 CC region. IL-D80 probably has either stimulatory or inhibitory  
 CC effects on haematopoietic cells, including lymphoid cells such as  
 CC T-cells, B-cells, natural killer cells, macrophages, dendritic  
 CC cells, haematopoietic progenitors, etc. It may be useful in the  
 CC treatment of immune disorders, such as T-cell immune deficiency,  
 CC chronic inflammation or tissue rejection, or in cardiovascular or  
 CC neurophysiological conditions. The full-length cytokine, and its  
 CC fragments or antagonists, will be useful in physiological modulation  
 CC of cells expressing a receptor. The proteins will also be useful as  
 CC antigens, e.g. immunogens, for raising antibodies to various  
 CC epitopes on the protein. Human and murine IL-D80 polypeptides and  
 CC polynucleotides are claimed, as well as expression vectors, host  
 CC cells, antigenic polypeptides, a detection kit, and a method of  
 CC modulating the physiology or development of a cell or tissue  
 CC culture cell by contacting with an agonist or antagonist of  
 CC primate IL-D80.  
 XX  
 XX Sequence 243 AA;  
 SQ

Query Match 0.6%; Score 8; DB 22; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 Db 163 peeeeeee 170  
 |||||

RESULT 44  
 AAG16482  
 ID AAG16482 standard; Protein; 254 AA.  
 XX  
 AC AAG16482;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 17145.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 07-MAY-1999; 99US-0132486.  
 PR 11-MAY-1999; 99US-0132487.  
 PR 14-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 19-MAY-1999; 99US-0134768.  
 PR 20-MAY-1999; 99US-0134941.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135353.  
 PR 25-MAY-1999; 99US-0135629.  
 PR 27-MAY-1999; 99US-0136021.  
 PR 28-MAY-1999; 99US-0136392.  
 PR 01-JUN-1999; 99US-0136782.  
 PR 03-JUN-1999; 99US-0137222.  
 PR 04-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137502.  
 PR 08-JUN-1999; 99US-0137724.  
 PR 10-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 14-JUN-1999; 99US-0138847.  
 PR 16-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 28-JUN-1999; 99US-0140695.  
 PR 29-JUN-1999; 99US-0140823.  
 PR 30-JUN-1999; 99US-0140991.  
 PR 01-JUL-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142154.  
 PR 06-JUL-1999; 99US-0142055.  
 PR 08-JUL-1999; 99US-0142390.  
 PR 09-JUL-1999; 99US-0142803.  
 PR 12-JUL-1999; 99US-0142920.  
 PR 13-JUL-1999; 99US-0142977.  
 PR 14-JUL-1999; 99US-0143542.  
 PR 15-JUL-1999; 99US-0143624.  
 PR 16-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 19-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.



Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
Db 169 eeeeeee 176  
|||||

RESULT 42  
AAB20275  
ID AAB20275 standard; Protein; 242 AA.  
XX AC AAB20275;  
XX DT 14-MAY-2001 (first entry)  
XX DE Human interleukin DNAX 80.  
XX KW Interleukin DNAX 80; IL-D80; human; cytokine; immunomodulator.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Peptide 1..27  
FT /label= Signal\_peptide  
FT Protein 28..242  
FT /label= Mature\_protein  
FT Region 35..56  
FT /label= Helix-A  
FT Region 87..113  
FT /label= Helix-B  
FT Region 123..156  
FT /label= Helix-C  
FT Region 203..230  
FT /label= Helix-D  
FT Region 13..20  
FT /note= "leucine repeat region"  
FT Region 163..175  
FT /note= "glutamic acid repeat region"  
FT Region 220..223  
FT /note= "leucine repeat region"  
XX XX  
XX WO200109176-A2.  
XX PN  
XX PD  
XX PF  
XX PR 30-JUL-1999; 99US-0364674.  
XX PR 06-AUG-1999; 99US-0369643.  
XX PA (SCHE ) SCHERING CORP.  
XX PI Timans JC, Kastelein RA, Bazan JF;  
XX DR WPI: 2001-182940/18.  
XX DR N-PSDB; AAF30407.  
XX XX  
XX PT New isolated or recombinant polynucleotide encoding an antigenic  
XX PT polypeptide (cytokine IL-D80) for providing reagents to regulate  
XX PT activation, development, differentiation and function of cells  
XX PT including haematopoietic cells  
XX XX  
XX PS Claim 1; Page 8; 73pp; English.  
XX XX  
XX CC The present sequence is that of a novel human cytokine, termed  
XX CC interleukin DNAX 80, or IL-D80, which shows sequence homology to  
XX CC interleukin-11. The sequence was deduced from isolated IL-D80  
XX CC cDNA (see AAF30407). A variant huIL-D80 sequence is provided in  
XX CC AAB20277. The 2 sequences differ only in the signal peptide  
XX CC region. IL-D80 probably has either stimulatory or inhibitory  
XX CC effects on haematopoietic cells, including lymphoid cells such as  
XX CC T-cells, B-cells, natural killer cells, macrophages, dendritic

CC cells, haematopoietic progenitors, etc. It may be useful in the  
CC treatment of immune disorders, such as T-cell immune deficiency,  
CC chronic inflammation or tissue rejection, or in cardiovascular or  
CC neurophysiological conditions. The full-length cytokine, and its  
CC fragments or antagonists, will be useful in physiological modulation  
CC of cells expressing a receptor. The proteins will also be useful as  
CC antigens, e.g. immunogens, for raising antibodies to various  
CC epitopes on the protein. Human and murine IL-D80 polypeptides and  
CC polynucleotides are claimed, as well as expression vectors, host  
CC cells, antigenic polypeptides, a detection kit, and a method of  
CC modulating the physiology or development of a cell or tissue  
CC culture cell by contacting with an agonist or antagonist of  
CC primate IL-D80.  
XX XX  
XX SQ Sequence 242 AA;

Query Match 0.6%; Score 8; DB 22; Length 242;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
Db 162 peeeeeee 169  
|||||

RESULT 43  
AAB20277  
ID AAB20277 standard; Protein; 243 AA.  
XX AC AAB20277;  
XX DT 14-MAY-2001 (first entry)  
XX DE Human interleukin DNAX 80 variant.  
XX KW Interleukin DNAX 80; IL-D80; human; cytokine; immunomodulator.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Peptide 1..28  
FT /label= Signal\_peptide  
FT Protein 29..243  
FT /label= Mature\_protein  
FT Region 35..57  
FT /label= Helix-A  
FT Region 89..114  
FT /label= Helix-B  
FT Region 124..157  
FT /label= Helix-C  
FT Region 204..231  
FT /label= Helix-D  
FT Region 14..21  
FT /note= "leucine repeat region"  
FT Region 164..176  
FT /note= "glutamic acid repeat region"  
FT Region 221..224  
FT /note= "leucine repeat region"  
XX XX  
XX WO200109176-A2.  
XX PN  
XX PD  
XX PF  
XX PR 27-JUL-2000; 2000WO-US20475.  
XX PR 30-JUL-1999; 99US-0364674.  
XX PR 06-AUG-1999; 99US-0369643.  
XX XX  
XX PA (SCHE ) SCHERING CORP.  
XX PI Timans JC, Kastelein RA, Bazan JF;  
XX XX



PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 EEEEEER 1332  
 DB 4 eeeeeer 11

RESULT 40  
 AAR31229  
 ID AAR31229 standard; protein; 242 AA.  
 AC AAR31229;

DT 19-MAY-1993 (first entry)  
 DE Prepro-thyrotropin releasing hormone.  
 XX Prepro TRH; antibodies; human; pituitary dysfunction; psychoses;  
 KW thyroidal diseases; thyroid.  
 XX Homo sapiens.

OS JP04352797-A.  
 PN 07-DEC-1992.  
 XX 29-MAY-1991; 91JP-0228139.  
 PF 29-MAY-1991; 91JP-0228139.  
 PR (SAKA ) OTSUKA PHARM CO LTD.  
 PA WPI; 1993-024099/03.  
 DR Human prepro-TRH related peptide(s) - used to prepare antibodies  
 PT which react specifically with human prepro-TRH for diagnosis of

PT e.g. pituitary dysfunction etc.  
 XX Disclosure; Page 14; 19pp; Japanese.  
 XX The sequence shown is that of human prepro-thyrotropin releasing hormone (TRH). The protein may be used to prepare related peptides which can be used to raise antibodies which specifically react with human prepro-TRH. The resultant antibodies may be used for the diagnosis of human TRH related diseases, e.g. pituitary dysfunction, thyroidal diseases and psychoses.  
 CC See also AAR31230-6.  
 XX Sequence 242 AA;

Query Match 0.6%; Score 8; DB 14; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 EEEEEER 1332  
 DB 169 eeeeeer 176

RESULT 41  
 AAW68538  
 ID AAW68538 standard; Protein; 242 AA.

AC AAW68538;  
 XX 12-FEB-1999 (first entry)  
 DE Pepto TRH protein, a TRAP molecule.  
 KW Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;  
 KW lymphoma; cancer; HLA; human lymphocyte antigen; vaccine.  
 XX Homo sapiens.

PN WO9849299-A1.  
 PD 05-NOV-1998.  
 XX 22-APR-1998; 98WO-US07784.  
 XX 25-APR-1997; 97US-0845998.

XX (LUDW-) LUDWIG INST CANCER RES.  
 XX Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;  
 DR WPI; 1999-009425/01.  
 XX N-PSDB; AAV33909.

PT New diagnosis of leukaemia - by detecting genes for tumour antigen  
 PT rejection precursors or corresponding proteins  
 XX Claim 11; Pages 54-55; 88pp; English.

XX The present sequence represents a tumour rejection antigen precursor (TRAP) designated TRH. The specification describes the treatment of disorders which characterised by expression of a leukaemia-associated nucleic acid such as TRH. The products are used for in vivo or in vitro screening for leukaemia, lymphoma or other cancers by usual hybridisation/amplification or immunoassay methods. TRAPS when processed to antigens or complexed with HLA (human lymphocyte antigen) molecules, or nucleic acid encoding them, are useful in vaccines for treating leukaemia.

XX Sequence 242 AA;

Query Match 0.6%; Score 8; DB 20; Length 242;



PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.



PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149502.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 237;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
Db 4 eeeeeer 11

## RESULT 39

AAG37419  
ID AAG37419 standard; Protein; 237 AA.

XX  
AC AAG37419;

XX  
DT 18-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46006.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX



```

AC AAW40075;
XX
XX 29-MAY-1998 (first entry)
XX
XX Guinea pig eosinophil granule MBP GMBP-1.
XX
XX Eosinophil-derived basic protein; EBPH; diagnosis; prevention; treatment;
XX KW disease; eosinophil accumulation; granule release; allergic reaction;
XX KW inflammatory reaction; eosinophilias; parasitic infection; pregnancy;
XX KW cytolytic agent; bacterial infection; cancer; major basic protein;
XX MBP; guinea pig.
XX
XX Cavia sp.
XX
XX US5728820-A.
XX
XX 17-MAR-1998.
XX
XX 23-OCT-1996; 96US-0740036.
XX
XX 23-OCT-1996; 96US-0740036.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Akerblom IE;
XX
XX WPI; 1998-206622/18.
XX
XX New DNA encoding eosinophil-derived basic protein - useful for
XX PT producing recombinant protein used for treating, e.g. parasitic
XX PT infections
XX
XX Disclosure; Fig 2; 26pp; English.
XX
XX This sequence represents the guinea pig eosinophil granule major basic
XX CC protein (MBP), GMBP-1, which is used to identify an eosinophil-derived
XX CC basic protein (EBPH) derived from interleukin-5 (IL-5) cultured human
XX CC umbilical cord blood cells. This protein can be used in the diagnosis,
XX CC prevention and treatment of diseases and conditions associated with
XX CC eosinophil accumulation and granule release, including late phase
XX CC allergic/inflammatory reactions, eosinophilias, parasitic infections and
XX CC conditions associated with placental-derived, eosinophil-derived basic
XX CC protein accumulation in pregnancy. The protein can also be used for
XX CC producing recombinant EBPH, which may be useful as a cytolytic agent in
XX CC the treatment of e.g. bacterial and parasitic infections and cancer.
XX
XX Sequence 233 AA;
XX

Query Match 0.6%; Score 8; DB 19; Length 233;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
Db 73 peeeeeee 80
|||||||

RESULT 38
AAG16483
ID AAG16483 standard; Protein; 237 AA.
XX
XX AAG16483;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 17146.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX
XX Arabidopsis thaliana.
XX OS

```

```

XX
XX PN EPI033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 03-JUN-1999; 99US-0137222.
XX 04-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.

```



PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 165;  
Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1324 PEEEEEE 1331  
Db 119 peeeeee 126  
RESULT 36  
AAR29716  
ID AAR29716 standard; Protein; 232 AA.  
XX  
AC AAR29716;  
XX  
DT 30-APR-1993 (first entry)  
XX  
DE Guinea pig MBP-1.  
XX  
KW Major basic protein; anti-asthma.  
XX  
OS Cavia porcellus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..15  
FT Peptide /note= "signal peptide"  
FT Peptide 16..113  
FT Protein /note= "prepeptide"  
FT Protein 114..232  
FT Protein /note= "mature MBP-1"  
XX  
PN JP04320686-A.  
XX  
PD 11-NOV-1992.  
XX  
PF 22-APR-1991; 91JP-0090704.  
XX  
PR 22-APR-1991; 91JP-0090704.  
XX  
PA (SAKA ) OTSUKA PHARM CO LTD.  
XX  
DR WPI; 1992-426680/52.  
DR N-PSDB; AAQ31988.  
XX  
PT A guinea pig MBP gene - contains DNA base sequence coding MBP  
PT gene having 1 of 2 specified aminoacid sequences  
XX  
PS Disclosure; Page 15; 20pp; Japanese.  
XX  
CC The sequence shows the protein sequence of the guinea pig major basic  
CC protein (MBP)-1. The sequence was obtd. by first isolating MBP-1  
CC and MBP-2 from the abdominal transudate of male guinea pigs treated  
CC with polmyxin and ascaris extract. MBP-1 and MBP-2 were used to  
CC design probes for screening of a guinea pig eosinophil cDNA library.  
CC The cDNA encoding the MBPs was used to transform suitable host  
CC cells which may then be cultured to obtain large amts. of MBP.  
CC The MBPs may be used for the confirmation of the effect of an anti-  
CC asthma drug. See also AAR29717.  
XX  
SQ Sequence 232 AA;  
Query Match 0.6%; Score 8; DB 13; Length 232;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1324 PEEEEEE 1331  
Db 72 peeeeee 79  
RESULT 37  
AAW40075  
ID AAW40075 standard; Protein; 233 AA.  
XX



PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 159;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
|  
Db 129 peeeeeee 136

## RESULT 35

AAG39616  
ID AAG39616 standard; Protein: 165 AA.

XX AC AAG39616;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49044.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 29-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.



PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149903.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.



PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	27-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151338.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
PR	21-OCT-1999;	99US-0160741.			
PR	21-OCT-1999;	99US-0160767.			
PR	21-OCT-1999;	99US-0160768.			
PR	21-OCT-1999;	99US-0160770.			
PR	21-OCT-1999;	99US-0160814.			
PR	21-OCT-1999;	99US-0160815.			
PR	22-OCT-1999;	99US-0160980.			
PR	22-OCT-1999;	99US-0160981.			
PR	23-JUL-1999;	99US-014522			



PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162143.

Query Match 0.6%; Score 8; DB 21; Length 149;  
Best Local Similarity 100.0%; Pred. No. 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

DB 119 peeeeeee 126

#### RESULT 33

AAG31670  
ID AAG31670 standard; Protein; 158 AA.

XX AC AAG31670;

XX DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38072.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0125264.

PR 29-MAR-1999; 99US-0125785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.



PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.



PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 149;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331.  
Db 119 peeeeee 126  
|||||

RESULT 32  
AAG37861  
ID AAG37861 standard; Protein; 149 AA.  
XX  
AC AAG37861;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46625.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.







PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 18-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 23-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.



PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 18-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 81;

Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||||

Db 51 peeeeee 58

RESULT 30

AAG39617

ID AAG39617 standard; Protein; 97 AA.

XX AC AAG39617;

XX AC AAG39617;

DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49045.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.

XX DE Arabidopsis thaliana.



PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

Db 51 peeeeee 58

RESULT 29

AAG37862  
 ID AAG37862 standard; Protein; 81 AA.

XX AC AAG37862;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45626.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135253.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.



KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 23-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134236.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

99US-0140354.

99US-0140695.

99US-0140823.

99US-0140891.

99US-0141287.

99US-0141842.

99US-0142154.

99US-0142055.

99US-0142390.

99US-0142803.

99US-0142920.

99US-0142977.

99US-0143542.

99US-0143624.

99US-0144005.

99US-0144085.

99US-0144086.

99US-0144325.

99US-0144331.

99US-0144332.

99US-0144333.

99US-0144334.

99US-0144335.

99US-0144352.

99US-0144632.

99US-0144884.

99US-0144814.

99US-0145086.

99US-0145088.

99US-0145085.

99US-0145087.

99US-0145089.

99US-0145192.

99US-0145145.

99US-0145218.

99US-0145224.

99US-0145276.

99US-0145913.

99US-0145918.

99US-0145919.

99US-0145951.

99US-0146386.

99US-0146389.

99US-0146388.

99US-0147038.

99US-0147204.

99US-0147302.

99US-0147302.

99US-0147192.

99US-0147260.

99US-0147303.

99US-0147416.

99US-0147493.

99US-0147935.

99US-0148171.

99US-0148319.

99US-0148341.

99US-0148565.

99US-0148684.

99US-0149368.

99US-0149368.

99US-0149175.

99US-0149426.

99US-0149722.

99US-0149723.

99US-0149929.

99US-0149902.

99US-0149930.

99US-0149930.

99US-0150566.

99US-0150884.

99US-0151065.

99US-0151066.

99US-0151086.

99US-0151303.

99US-0151438.



PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1998; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1998; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 74;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 1325 EEEEEER 1332  
Db 42 eeeeeer 49

RESULT 28  
AAG19544  
ID AAG19544 standard; Protein; 81 AA.  
XX  
AC AAG19544;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21385.  
XX



```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 58;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 13 eeeeeer 20

RESULT 27
AAG41122
ID AAG41122 standard; Protein; 74 AA.
```

```
XX
AC AAG41122;
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 51122.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
```



[illegible]



PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 0.68; Score 8; DB 21; Length 58;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1325 EEEEEER 1332



SQ Sequence 28 AA;

Query Match 0.6%; Score 8; DB 14; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
| | | | | | | |  
Db 14 eeeeeer 21

RESULT 24

AAR95846  
ID AAR95846 standard; Protein; 32 AA.

XX  
AC AAR95846;

XX  
DT 28-OCT-1996 (first entry)

XX  
DE Human prepro-TRH peptide having CRIF activity (aa 152-183).

XX  
KW TRH; thyrotropin release factor; Cushing's disease; anxiety;  
KW corticotropin release inhibiting factor; CRIF; depression; obesity;  
KW anorexia nervosa; withdrawal; hypocortisolism; colitis; autoimmune;  
KW arthritis; premenstrual syndrome; inflammatory; obsessive compulsive;  
KW disorder.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT Modified-site 1

FT /label= OTHER

FT /note= "pyroglutamic acid"

FT Misc-difference 1..6

FT /note= "residues 1-6 may be opt. cleaved  
i.e. absent"

XX  
PN WO9608265-A1.

XX  
PD 21-MAR-1996.

XX  
PF 08-SEP-1995; 95WO-US11455.

XX  
PR 12-SEP-1994; 94US-0304383.

XX  
PA (UYPE-) UNIV PENNSYLVANIA.

XX  
PI Aird F, Redel E;

XX  
XX WPI; 1996-179720/18.

DR  
N-PSDB; AAT15289.

XX  
PT New isolated corticotropin release inhibiting factor peptide(s) -  
PT used to develop prods. for the diagnosis and treatment of CRIF  
PT related disorders, e.g. stress responses or inflammation

XX  
PS Claim 7; Figure 10; 66pp; English.

XX  
CC AAR95846 is a portion (amino acids 152-183) of the human prepro-TRH  
CC (thyrotropin releasing hormone) protein positioned between the  
CC fourth and fifth TRH sequences. The peptide has CRIF (corticotropin  
CC release inhibiting factor) activity and may be used for the  
CC diagnosis and treatment of CRIF related disorders. Such disorders  
CC include Cushing's disease, anxiety, anorexia nervosa, depression,  
CC obesity, withdrawal from drug or alcoholic dependence, some cancers,  
CC hypercortisolism, ACTH (adrenocorticotropin) deficiency; premenstrual  
CC syndrome, inflammatory conditions e.g. colitis and autoimmune disease  
CC e.g. arthritis. Peptides with CRIF activity may also be used to  
CC enhance immune responses and to increase the efficacy of immunity to  
CC vaccines.

XX  
SQ Sequence 32 AA;

Query Match 0.6%; Score 8; DB 17; Length 32;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
| | | | | | | |  
Db 18 eeeeeer 25

RESULT 25

AAG56078

ID AAG56078 standard; Protein; 58 AA.

XX  
AC AAG56078;

XX  
DT 18-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 72021.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX  
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0128845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.



KW post-traumatic stress disorder; therapy.

XX Homo sapiens.

XX US6039956-A.

XX 21-MAR-2000.

XX 07-JUN-1996; 96US-0660561.

XX 12-SEP-1994; 94US-0304383.

XX 08-SEP-1995; 95US-0523125.

XX (UYPE-) UNIV PENNSYLVANIA.

XX McGivern RF, Rittenhouse PA, Aird F, Redei E;

XX WPI; 2000-270120/23.

XX N-PSDB; AAA07229.

XX Treating behavioural symptoms including panic disorder, post-traumatic stress disorder and obsessive compulsive disorder in humans, involves administering corticotropin release inhibiting factor -

XX Claim 5; Column 35; 43pp; English.

XX This sequence is a human corticotropin release inhibitory factor (CRIF) fragment. The invention relates to a method for treating (I) behavioural symptoms in an anxiety disorder in a human, comprising administering CRIF which comprises 3 contiguous amino acids contained within the amino acid sequence positioned between the fourth and fifth thyrotropin releasing hormone (TRH) sequence on a prepro-TRH protein. (I) is useful for treating behavioural symptoms in an anxiety disorder which includes general anxiety disorder, panic disorder, obsessive compulsive disorder and post-traumatic stress disorder. Peripheral administration of CRIF is capable of crossing the blood-brain barrier.

XX Sequence 26 AA;

Query Match 0.6%; Score 8; DB 21; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 12 eeeeeer 19

RESULT 22

AAB61357

ID AAB61357 standard; peptide; 26 AA.

XX AAB61357;

XX 03-APR-2001 (first entry)

XX Human corticotropin release inhibiting factor peptide.

XX Depressive disorder; corticotropin release inhibiting factor; CRIF; bipolar; dysthymia; cyclothymia; premenstrual.

XX Homo sapiens.

XX WO200100222-A1.

XX 04-JAN-2001.

XX 09-JUN-2000; 2000WO-US16107.

XX 24-JUN-1999; 99US-0140962.

XX 04-AUG-1999; 99US-0366981.

XX

PA (NOUN ) UNIV NORTHWESTERN.

XX Redei E;

XX WPI; 2001-091671/10.

XX Treating depressive disorder such as major and minor depression, bipolar disorders, dysthymia, cyclothymia, and premenstrual syndrome, involves administering a compound having corticotropin release inhibiting factor activity -

XX Claim 15; Page 28; 36pp; English.

XX The present invention relates to treating a depressive disorder in an animal, involving administering a compound with corticotropin release inhibiting factor (CRIF) activity. The invention is useful for treating depressive disorders such as major depression, minor depression, bipolar disorders, dysthymia, cyclothymia and premenstrual syndrome, in human.

XX Sequence 26 AA;

Query Match 0.6%; Score 8; DB 22; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 12 eeeeeer 19

RESULT 23

AAR31234

ID AAR31234 standard; peptide; 28 AA.

XX AAR31234;

XX 19-MAY-1993 (first entry)

XX Prepro-thyrotropin releasing hormone related peptide.

XX Prepro TRH; antibodies; human; pituitary dysfunction; psychoses; thyroidal diseases; thyroid.

XX Synthetic.

XX JF04352797-A.

XX 07-DEC-1992.

XX 29-MAY-1991; 91JP-0228139.

XX 29-MAY-1991; 91JP-0228139.

XX (SAKA ) OTSUKA PHARM CO LTD.

XX WPI; 1993-024099/03.

XX Human prepro-TRH related peptide(s) - used to prepare antibodies which react specifically with human prepro-TRH for diagnosis of e.g. pituitary dysfunction etc.

XX Claim 5; Page 16; 19pp; Japanese.

XX The sequence shown is that of a fragment of human prepro-thyrotropin releasing hormone (TRH) which may be prep. synthetically. The peptide can be used to raise antibodies which specifically react with human prepro-TRH. The resultant antibodies may be used for the diagnosis of human TRH related diseases, e.g. pituitary dysfunction, thyroidal diseases and psychoses.

XX See also AAR31229-36.

XX



```

SQ Sequence 1886 AA;
Query Match 0.6%; Score 9; DB 19; Length 1886;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKOLEV 565
Db 1476 ekirkqlev 1484
|||||

RESULT 19
AAR31233
ID AAR31233 standard; peptide; 26 AA.
XX
AC AAR31233;
XX
DT 19-MAY-1993 (first entry)
XX
XX Prepro-thyrotropin releasing hormone related peptide.
DE XX
XX Prepro TRH; antibodies; human; pituitary dysfunction; psychoses;
KW KW thyroidal diseases; thyroid.
XX
OS Synthetic.
XX
XX JP04352797-A.
PN PN
PD PD
XX 07-DEC-1992.
XX
PF 29-MAY-1991; 91JP-0228139.
XX
PR 29-MAY-1991; 91JP-0228139.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
PA
XX
XX WPI; 1993-024099/03.
DR XX
XX Human prepro-TRH related peptide(s) - used to prepare antibodies
PT which react specifically with human prepro-TRH for diagnosis of
PT e.g. pituitary dysfunction etc.
XX
PS Claim 4; Page 16; 19pp; Japanese.
XX
XX The sequence shown is that of a fragment of human prepro-thyrotropin
CC releasing hormone (TRH) which may be prep. synthetically. The
CC peptide can be used to raise antibodies which specifically react with
CC human prepro-TRH. The resultant antibodies may be used for the
CC diagnosis of human TRH related diseases, e.g. pituitary dysfunction,
CC thyroidal diseases and psychoses.
CC See also AAR31229-36.
XX
SQ Sequence 26 AA;

Query Match 0.6%; Score 8; DB 14; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 12 eeeeeer 19
|||||

RESULT 20
AAW73133
ID AAW73133 standard; peptide; 26 AA.
XX
AC AAW73133;
XX
XX 15-JAN-1999 (first entry)
DT
KW
KW

```

---

```

DE XX
KW Corticotropin release inhibiting factor peptide.
KW CRIF; corticotropin release inhibiting factor; Cushing's disease; cancer;
KW hypothalamic corticotropin-releasing factor; anorexia nervosa; obesity;
KW pituitary adrenocorticotropin; adrenocortical glucocorticoid; depression;
KW thyrotropin-releasing hormone; CRIF disorder; anxiety; viral infection;
KW therapy.
XX
XX Homo sapiens.
XX OS
XX US5830866-A.
XX PN
XX 03-NOV-1998.
XX PD
XX 08-SEP-1995; 95US-0523125.
XX PF
XX 08-SEP-1995; 95US-0523125.
XX PR
XX 12-SEP-1994; 94US-0304383.
XX PR
XX (UYPE-) UNIV PENNSYLVANIA.
XX PA
XX Aird F, Redei E;
XX PI
XX WPI; 1998-609292/51.
XX DR
XX Reduction of e.g. pituitary adrenocorticotropin - comprises
PT administration of fragment of prepro thyrotropin-releasing hormone
PT protein
XX
XX Claim 7; Column 26; 33pp; English.
XX
XX This sequence represents a corticotropin release inhibiting factor (CRIF)
CC peptide. This CRIF peptide can be used in the method of the invention for
CC reducing the level of hypothalamic corticotropin-releasing factor,
CC pituitary adrenocorticotropin and/or adrenocortical glucocorticoid
CC comprises administration of a CRIF suspended in a carrier. CRIF is a
CC recombinant or synthetic peptide comprising at least three contiguous
CC amino acids from the amino acid sequence positioned between the fourth
CC and fifth thyrotropin-releasing hormone (TRH) sequences of a prepro-TRH
CC protein. The method is used for the treatment of CRIF disorders
CC characterised by elevated levels of hypothalamic corticotropin-releasing
CC factor, pituitary adrenocorticotropin and/or adrenocortical
CC glucocorticoid. Such disorders include Cushing's disease,
CC anxiety, anorexia nervosa, depression, obesity, alcohol and drug
CC withdrawal symptoms, disorders exacerbated by stress-induced
CC immunosuppression (e.g. viral infections), and cancers whose severity is
CC reduced by an enhanced immune response.
XX
SQ Sequence 26 AA;

Query Match 0.6%; Score 8; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 12 eeeeeer 19
|||||

RESULT 21
AAAY81900
ID AAAY81900 standard; Peptide; 26 AA.
XX
XX AC AAAY81900;
XX
XX 22-JUN-2000 (first entry)
DT
XX Human CRIF protein sequence fragment.
DE
XX Corticotropin release inhibitory factor; CRIF; behavioural symptom;
KW anxiety disorder; thyrotropin releasing hormone; TRH sequence; human;
KW general anxiety disorder; panic disorder; obsessive compulsive disorder;

```



Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 17  
AAB01832  
ID AAB01832 standard; Protein; 1011 AA.  
XX AC AAB01832;  
XX DT 11-SEP-2000 (first entry)  
XX DE Haemophilus influenzae strain K21  
XX HW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;  
KW recombinant production; Escherichia coli; antibacterial; vaccine;  
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
KW detection; diagnosis.  
XX DE Haemophilus influenzae strain K21.  
XX PN WO200020609-A2.  
XX PD 13-APR-2000.  
XX PF 07-OCT-1999; 99WO-CA00938.  
XX PR 07-OCT-1998; 98US-0167568.  
XX PR 08-DEC-1998; 98US-0206942.  
XX PA (CONN-) CONNAUGHT LAB LTD.  
XX PI Loosmore SM, Yang Y, Klein MH;  
XX DR WPI: 2000-303789/26.  
XX DR N-PSDB; AAA52181.  
XX PT Nucleic acid molecule for producing recombinant high molecular weight  
PT proteins of Haemophilus which are used as a vaccine to provide  
PT protection against Haemophilus induced diseases in humans -  
XX PS Claim 12; Fig 21A-O; 307pp; English.  
XX CC The invention relates to the recombinant production of Haemophilus  
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
CC expression construct used to effect recombinant expression comprises a  
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene  
CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,  
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins  
CC and the hmwB and hmwC genes encode accessory proteins which are  
CC responsible for post-translational processing and secretion of the HMWA  
CC proteins. The modified hmwABC operon used in the expression construct of  
CC the invention contains an A gene modified such that it encodes only the  
CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)  
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae  
CC strains Joyce, K1, K21, LDCD2, PWH1, 15 and 12. The nucleic acids and  
CC vectors are used for the production of recombinant H. influenzae HMW  
CC proteins which can be used as vaccines to mediate a humoral or  
CC cell-mediated immune response to provide protection against diseases in  
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
CC antigens in immunoassays for detecting antibodies against Haemophilus,  
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
CC HMW proteins can be used to isolate and clone hmw genes from other  
CC non-typeable strains of Haemophilus via hybridisation reactions. The  
CC present sequence represents an HMWA protein from a non-typeable strain of  
CC H. influenzae.  
XX SQ Sequence 1011 AA;

Query Match 0.6%; Score 9; DB 21; Length 1011;

QY 278 EFKLKSDN 286  
| | | | | | | |

Db 269 efklksdn 277

RESULT 18

AAW54241  
ID AAW54241 standard; Protein; 1886 AA.

XX AC AAW54241;

XX DT 17-AUG-1998 (first entry)

XX DE Rattus norvegicus mutant alpha-myosin heavy chain.

XX Actin binding domain; alpha myosin heavy chain; bridging; mouse;  
KW transgenic mammal; congestive heart failure; study; treatment; diet;  
KW exercise; effects; identification; hypertrophic cardio-myopathy;  
KW dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;  
KW tricuspid stenosis; constrictive pericarditis; hypertension;  
KW acute infective endocarditis; ischaemic heart disease;  
KW primary myocardial disease; valvular disease; pericardial disease;  
KW hyperthyroidism; anaemia; arteriovenous fistula; beri-beri;  
KW Paget's disease; transgene.

XX Rattus norvegicus.

OS Synthetic.

XX FH Key Location/Qualifiers

XX FT Region 403

XX FT /note= "Arg403Gln mutation"

XX FT Region 468..475

XX FT /note= "nonmyosin actin binding domain bridge"

XX WO9813476-A1.

XX PD 02-APR-1998.

XX PF 26-SEP-1997; 97WO-US17296.

XX PR 26-SEP-1996; 96US-0026742.

XX PA (LEIN/) LEINWAND LA.

XX PI Vikstrom KL;

XX DR WPI; 1998-230690/20.

XX DR N-PSDB; AAV21518.

XX Transgenic mouse models for congestive heart failure and  
PT hypertrophic cardio-myopathy - used to study molecular and cellular  
PT events, identify potential therapeutic agents, assess effects of  
PT diet etc.

XX PS Disclosure; Pages 53-58; 75pp; English.

XX The sequence is that of a mutant rat alpha-myosin heavy chain  
CC which was used in the development of transgenic mammals,  
CC specifically mice. They can be used as a model for studying congestive  
CC heart failure (CHF) or hypertrophic cardiomyopathy. Such animals  
CC are used to study molecular and cellular events associated with  
CC CHF; to identify compounds for treating CHF, and in evaluating  
CC effects of diet and exercise on CHF. Conditions associated with CHF  
CC that can be evaluated this way are dilated or hypertrophic  
CC cardiomyopathy; acute aortic regurgitation; tricuspid stenosis;  
CC constrictive pericarditis; acute infective endocarditis; ischaemic  
CC heart disease; hypertension; primary myocardial disease; valvular  
CC or pericardial disease; hyperthyroidism; anaemia; arteriovenous  
CC fistula; beri-beri and Paget's disease.



```

SQ Sequence 439 AA;
Query Match 0.6%; Score 9; DB 21; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1331
Db 53 apeeeeeee 61
|||||

RESULT 15
AAB66298
ID AAB66298 standard; Protein; 439 AA.
XX
AC AAB66298;
XX
DT 05-APR-2001. (first entry)
XX
DE Human TRF1 SEQ ID NO: 151.
XX
KW Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasm; aging;
KW inflammatory disorder.
XX
OS Homo sapiens.
XX
PN WO200100849-A1.
XX
PD 04-JAN-2001.
XX
PF 28-JUN-2000; 2000WO-US17827.
XX
PR 29-JUN-1999; 99US-0141582.
XX
PA (ICOS-) ICOS CORP.
XX
PI Christenson E, Demaggio AJ, Goldman PS, McElligott DL;
XX
WPI; 2001-102896/11.
DR N-PSDB; AAF63966.
XX
XX New tankyrase2 polypeptides, useful for treating conditions mediated by
PT poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,
PT inflammatory and autoimmune disorders -
XX
XX Example 6; Page 223-225; 242pp; English.
XX
CC The present invention provides the protein and coding sequence for the
CC human tankyrase2 protein. This is found in two different versions,
CC designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has
CC polyADP-ribosylation activity and is involved in the modification of
CC TRF1, which is a telomere-specific binding protein. The regulation of
CC telomere length, in which TRF1 has a role, is linked to ageing and
CC cancer. The sequences are useful in the treatment of cancers and
CC inflammatory disorders.
XX
SQ Sequence 439 AA;
Query Match 0.6%; Score 9; DB 22; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1331
Db 53 apeeeeeee 61
|||||

RESULT 16
AAB01833
ID AAB01833 standard; Protein; 1005 AA.
XX
AC AAB01833;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain K21.
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX
PR 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
WPI; 2000-303789/26.
DR N-PSDB; AAA52182.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX Claim 8; Fig 21A-O; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAA521/5-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains JOYC, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typeable
CC strain of H. influenzae.
XX
SQ Sequence 1005 AA;
Query Match 0.6%; Score 9; DB 21; Length 1005;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 EFKLKSNNDN 286
Db 263 efklksndn 271
|||||

```



PS Disclosure; Page 112-114; 163pp; English.

XX The telomere repeat binding factor (TRF) is used in the production of an  
 CC altered vertebrate telomere repeat binding protein (A-TRF) which has a  
 CC TRF dimerisation domain, and forms a hetero-dimer with TRF, preventing it  
 CC from binding to the specified repeat sequence. A-TRF, optionally  
 CC expressed by gene therapy, is used to inhibit shortening of telomeres  
 CC associated with ageing (for cosmetic purposes) and disease, e.g. ataxia  
 CC telangeiectasia, Down's syndrome, atrophy of the skin, age-related macular  
 CC degeneration, atherosclerosis, tumours and viral (including human immune  
 CC deficiency virus) infection. Cells expressing A-TRF also have an  
 CC increased life span in vitro, e.g. for expression of recombinant proteins  
 CC or where intended for subsequent transplant or for testing, eliminating  
 CC the need for transformation.

XX Sequence 439 AA;

Query Match 0.6%; Score 9; DB 19; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331  
 Db 53 apeeeeeee 61  
 |||||

RESULT 13

AAAY02182  
 ID AAY02182 standard; Protein; 439 AA.

XX

AC AAY02182;

XX 07-JUL-1999 (first entry)

XX A human telomere repeat binding factor (TRF).

XX Telomere repeat binding factor; TRF; TRF2; telomere termini;  
 KW cell proliferation; antagonist; agonist; telomere length reduction;  
 KW aging; skin atrophy; age-related macular degeneration; atherosclerosis;  
 KW abnormal telomere length; cancer cell; telomere-associated disorder;  
 KW mammalian artificial chromosome; gene therapy; viral infection.

XX Homo sapiens.

XX WO9915662-A1.

XX 01-APR-1999.

XX 25-SEP-1998; 98WO-US20175.

XX 04-FEB-1998; 98US-0018635.

XX 26-SEP-1997; 97US-0938052.

XX (UYRQ ) UNIV ROCKEFELLER.

XX Broccoli D, De Lange T, Smogorzenska A;  
 PI WPI; 1999-263481/22.

XX N-PSDB; AAX35386.

XX Nucleic acid encoding telomere repeat binding factor for treatment  
 PT of, e.g. cancer

XX Disclosure; Fig 3; 160pp; English.

XX The present sequence represents human telomere repeat binding  
 CC factor (TRF). The specification also describes vertebrate  
 CC TRF2 polynucleotides and polypeptides. When the basic N-terminal  
 CC domain of the TRF2 protein is removed, it binds detectably to the  
 CC telomere repeat sequence (TTAGGG)<sub>12</sub>. TRF maintains the correct  
 CC structure of telomere termini and protects against end-to-end fusion.  
 CC It is required for cell proliferation. TRFs (or their antagonists and

CC agonists) can be used to limit reduction in telomere length associated  
 CC with aging (e.g. atrophy of the skin, age-related macular degeneration  
 CC and atherosclerosis) or abnormal telomere lengths in cancer cells.  
 CC They can also be used to screen for specific modulators (potentially  
 CC useful for treating aging and cancer), including those that are specific  
 CC for one TRF over another; to diagnose telomere-associated disorders; as  
 CC targeting agents for TTAGG repeats and in construction of mammalian  
 CC artificial chromosomes (for gene therapy or basic research). Antibodies  
 CC against TRF can be used for diagnosis and therapy, e.g. to differentiate  
 CC between different TRFs to screen expression libraries for TRF-expressing  
 CC genes or to detect (pre)cancers or viral infections.

XX Sequence 439 AA;

Query Match 0.6%; Score 9; DB 20; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331  
 Db 53 apeeeeeee 61  
 |||||

RESULT 14

AAB03942

ID AAB03942 standard; Protein; 439 AA.

XX

AC AAB03942;

XX 26-FEB-2001 (first entry)

XX Human telomeric repeat binding factor 1.

XX Human telomeric repeat binding factor 1; hTRBF-1; antisense;  
 KW disease; modulation; expression; prophylaxis; infection;  
 KW inflammation; anti-inflammatory; tumour; diagnostic; human.

XX Homo sapiens.

XX US6130088-A.

XX 10-OCT-2000.

XX 21-JUL-1999; 99US-0358384.

XX 21-JUL-1999; 99US-0358384.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Cowser LM;

XX WPI; 2000-664192/64.

XX N-PSDB; AAA54235.

XX New antisense compounds that hybridizes with and inhibits the  
 PT expression of human telomeric repeat binding factor 1 (TRBF-1), useful  
 PT for treating conditions or diseases associated with TRBF-1 expression  
 XX Example 1; Columns 47-50; 34pp; English.

XX Antisense compounds directed against the start codon, 3' untranslated  
 CC region, nucleotides 78-374, 560-681, 965-1334 of the coding region or  
 CC the stop codon of the human telomeric repeat binding factor 1  
 CC (hTRBF-1) can be used to inhibit the expression of hTRBF-1. The  
 CC antisense compounds are used for treating a patient suspected of  
 CC having or being prone to a disease or condition associated with  
 CC expression of TRBF-1 by modulating its expression. They may also be  
 CC used prophylactically to prevent or delay infection, inflammation or  
 CC tumour formation, or as research reagents and diagnostics, e.g. to  
 CC distinguish between functions of various members of a biological  
 CC pathway.



DT 14-DEC-1998 (first entry)  
 XX  
 DE Altered telomere repeat binding factor protein.  
 XX  
 KW human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 XX telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
 XX Homo sapiens.  
 OS Synthetic.  
 OS  
 PN WO9836066-A1.  
 XX  
 XX 20-AUG-1998.  
 PD  
 XX 13-FEB-1998; 98WO-US02765.  
 PF  
 XX 04-FEB-1998; 98US-0018628.  
 PR  
 PR 13-FEB-1997; 97US-0800264.  
 XX  
 XX (UYRQ ) UNIV ROCKEFELLER.  
 PA  
 XX Bianchi A, De Lange T, Van Steensel B;  
 PI  
 XX WPI: 1998-480769/41.  
 DR N-PSDB; AAV59280.  
 DR  
 XX Nucleic acid encoding altered telomere repeat binding protein and  
 XX related vectors - transformants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 PT  
 XX Claim 13; Page 115-117; 163pp; English.  
 PS  
 XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
 CC inhibit shortening of telomeres associated with ageing (for cosmetic  
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
 CC tumours and viral (including human immune deficiency virus) infection.  
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
 CC for expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.  
 CC  
 XX Sequence 437 AA;  
 SQ  
 Query Match 0.6%; Score 9; DB 19; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1323 APEEEEEEE 1331  
 Db 53 apeeeeeee 61  
 RESULT 11  
 AAW28555  
 ID AAW28555 standard; Protein; 439 AA.  
 XX  
 AC AAW28555;  
 XX  
 XX 09-FEB-1998 (first entry)  
 DT  
 DE Telomeric repeat binding factor.  
 XX  
 XX Telomeric repeat binding factor; TRF; ageing; cancer.  
 KW  
 XX Homo sapiens.  
 OS  
 OS WO9708314-A2.  
 PN  
 XX

PD 06-MAR-1997.  
 XX  
 XX 23-AUG-1996; 96WO-US13652.  
 PF  
 XX 25-AUG-1995; 95US-0519103.  
 PR  
 XX (UYRQ ) UNIV ROCKEFELLER.  
 PA  
 XX De LANGE T;  
 PI  
 XX WPI: 1997-179271/16.  
 DR N-PSDB; AAT87489.  
 DR  
 XX Telomeric repeat binding factor of mol. wt. 67 kD - useful for  
 PT diagnosis and treatment of cellular ageing and cancer  
 PT  
 XX Claim 4; Fig 3; 62pp; English.  
 PS  
 XX TRF is the first telomeric protein isolated from a multicellular  
 CC organism. Human telomeres shorten during normal cell division  
 CC and this shortening may eventually limit cell proliferation  
 CC and lead to ageing. Addition of TRF, its analogues, agonists  
 CC or mimics will be useful for restoring telomere function to treat  
 CC cellular ageing.  
 CC  
 XX Sequence 439 AA;  
 SQ  
 Query Match 0.6%; Score 9; DB 18; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1323 APEEEEEEE 1331  
 Db 53 apeeeeeee 61  
 RESULT 12  
 AAW77348  
 ID AAW77348 standard; Protein; 439 AA.  
 XX  
 AC AAW77348;  
 XX  
 XX 14-DEC-1998 (first entry)  
 DT  
 DE Human telomere repeat binding factor protein.  
 XX  
 XX Human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO9836066-A1.  
 PN  
 XX 20-AUG-1998.  
 PD  
 XX 13-FEB-1998; 98WO-US02765.  
 PF  
 XX 04-FEB-1998; 98US-0018628.  
 PR  
 PR 13-FEB-1997; 97US-0800264.  
 XX  
 XX (UYRQ ) UNIV ROCKEFELLER.  
 PA  
 XX Bianchi A, De Lange T, Van Steensel B;  
 PI  
 XX WPI: 1998-480769/41.  
 DR N-PSDB; AAV59278.  
 DR  
 XX Nucleic acid encoding altered telomere repeat binding protein and  
 PT related vectors - transformants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 PT  
 XX



CC cancer. The sequences are useful in the treatment of cancers and  
 CC inflammatory disorders.

XX  
 SQ Sequence 67 AA;

Query Match 0.6%; Score 9; DB 22; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 0.37;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331  
 |||||

Db 53 apeeeeeee 61

RESULT 8  
 AAG04041  
 ID AAG04041 standard; Protein; 114 AA.

XX AC AAG04041;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 8122.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI; 2000-500381/45.

XX N-PSDB; AAC04047.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 8122; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 114 AA;

Query Match 0.6%; Score 9; DB 21; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 0.61;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

Db 53 apeeeeeee 61  
 |||||

RESULT 9

AAW77351

ID AAW77351 standard; Protein; 325 AA.

XX AC AAW77351;

XX 14-DEC-1998 (first entry)

XX Altered telomere repeat binding factor protein deletion mutant.

XX Human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

XX Homo sapiens.

XX Synthetic.

XX WO9836066-A1.

XX 20-AUG-1998.

XX 13-FEB-1998; 98WO-US02765.

XX 04-FEB-1998; 98US-0018628.

XX 13-FEB-1997; 97US-0800264.

XX (UYRQ ) UNIV ROCKEFELLER.

XX Bianchi A, De Lange T, Van Steensel B;

XX WPI; 1998-480769/41.

XX N-PSDB; AAV59281.

XX Nucleic acid encoding altered telomere repeat binding protein and  
 PT related vectors - transformants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture

XX Claim 15; Page 117-118; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
 CC inhibit shortening of telomerases associated with ageing (for cosmetic  
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
 CC tumours and viral (including human immune deficiency virus) infection.  
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
 CC for expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.

XX SQ Sequence 325 AA;

Query Match 0.6%; Score 9; DB 19; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

|||||

Db 53 apeeeeeee 61

RESULT 10

AAW77350

ID AAW77350 standard; Protein; 437 AA.

XX AC AAW77350;

XX



XX PS Claim 11; Page 1337-1339; 1425pp; English.

XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

XX CC associated proteins and polynucleotide sequences, their agonists, and

XX CC antagonists may have neuroprotective; cytostatic; cardioactive;

XX CC immunomodulatory; muscular active general; vulnary; gastrointestinal

XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial

XX CC activity. The invention also includes antibodies specific for the

XX CC protein or polynucleotide sequences. The lung cancer associated

XX CC polynucleotide sequences may be used for detection of lung cancer.

XX CC chromosome identification, as chromosome markers, and for numerous other

XX CC diagnostic or research purposes. The proteins may be used to treat

XX CC disorders such as neural, immune, muscular, reproductive,

XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

XX CC disorders. The proteins may also be used in the treatment of wounds and

XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

XX CC peptide AAB58549 are used in the course of the invention for the

XX CC identification and characterisation of the polynucleotide and protein

XX CC sequences.

XX SQ Sequence 363 AA;

Query Match 1.0%; Score 14; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 9.1e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 EKLISLLPEYVVPY 994

Db |||||||

17 eklisllpeyvvpv 30

RESULT 6

AAAY02188

ID AAY02188 standard; Protein: 67 AA.

XX AC AAY02188;

XX DT 07-JUL-1999 (first entry)

XX DE Polypeptide sequence from SEQ ID 33.

XX KW Telomere repeat binding factor; TRF; TRF2; telomere termini;

XX KW cell proliferation; antagonist; agonist; telomere length reduction;

XX KW aging; skin atrophy; age-related macular degeneration; atherosclerosis;

XX KW abnormal telomere length; cancer cell; telomere-associated disorder;

XX KW mammalian artificial chromosome; gene therapy; viral infection.

XX OS Homo sapiens.

XX PN WO9915662-A1.

XX PD 01-APR-1999.

XX PF 25-SEP-1998; 98WO-US20175.

XX PR 04-FEB-1998; 98US-0018635.

XX PR 26-SEP-1997; 97US-0938052.

XX PA (UYRQ ) UNIV ROCKEFELLER.

XX PI Broccoli D, De Lange T, Smogorzenska A;

XX DR WPI; 1999-263481/22.

XX DR N-PSDB; AAX35393.

XX XX Nucleic acid encoding telomere repeat binding factor for treatment

XX PT of, e.g. cancer

XX PS Disclosure; Page 133; 160pp; English.

XX XX

CC The specification describes vertebrate telomere repeat binding

CC factor (TRF) and TRF2 polynucleotides and polypeptides. When the basic

CC N-terminal domain of the TRF protein is removed, it binds detectably to

CC the telomere repeat sequence (TTAGGG)<sub>12</sub>. TRF maintains the correct

CC structure of telomere termini and protects against end-to-end fusion.

CC It is required for cell proliferation. TRFs (or their antagonists and

CC agonists) can be used to limit reduction in telomere length associated

CC with aging (e.g. atrophy of the skin, age-related macular degeneration

CC and atherosclerosis) or abnormal telomere lengths in cancer cells.

CC They can also be used to screen for specific modulators (potentially

CC useful for treating aging and cancer), including those that are specific

CC for one TRF over another: to diagnose telomere-associated disorders; as

CC targeting agents for TTAGGG repeats and in construction of mammalian

CC artificial chromosomes (for gene therapy or basic research). Antibodies

CC against TRF can be used for diagnosis and therapy, e.g. to differentiate

CC between different TRFs to screen expression libraries for TRF-expressing

CC genes or to detect (pre)cancers or viral infections. The present

CC sequence is used in the course of the invention.

XX SQ Sequence 67 AA;

Query Match 0.6%; Score 9; DB 20; Length 67;

Best Local Similarity 100.0%; Pred. No. 0.37;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEE 1331

Db |||||||

53 apeeeeeee 61

RESULT 7

AAAB66299

ID AAB66299 standard; Protein: 67 AA.

XX AC AAB66299;

XX DT 05-APR-2001 (first entry)

XX DE Human TRF1 TANK1 binding domain protein sequence SEQ ID NO: 155.

XX KW Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasm; aging;

XX KW inflammatory disorder.

XX OS Homo sapiens.

XX PN WO200100849-A1.

XX PD 04-JAN-2001.

XX PF 28-JUN-2000; 2000WO-US17827.

XX PR 29-JUN-1999; 99US-0141582.

XX PA (ICOS-) ICOS CORP.

XX PI Christenson E, Demaggio AJ, Goldman PS, McElligott DL;

XX DR WPI; 2001-102896/11.

XX DR N-PSDB; AAF63969.

XX PT New tankyrase2 polypeptides, useful for treating conditions mediated by

XX PT poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,

XX PT inflammatory and autoimmune disorders -

XX PS Example 6; Page 225-226; 242pp; English.

XX CC The present invention provides the protein and coding sequence for the

XX CC human tankyrase2 protein. This is found in two different versions,

XX CC designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has

XX CC polyADP-ribosylation activity and is involved in the modification of

XX CC TRF1, which is a telomere-specific binding protein. The regulation of

XX CC telomere length, in which TRF1 has a role, is linked to ageing and



FT Modified-site 492 "Potential glycosylation site"  
 FT Modified-site 494  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 507  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 557  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 561  
 FT Modified-site /note= "Potential glycosylation site"  
 FT Modified-site 582  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 593  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 594  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 614  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 626  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 649  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 668  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 677  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 697  
 FT Modified-site /note= "Potential glycosylation site"  
 FT Modified-site 712  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 747  
 FT Modified-site /note= "Potential glycosylation site"  
 FT Modified-site 750  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 757  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 798  
 FT Modified-site /note= "Potential glycosylation site"  
 FT Modified-site 800  
 FT Modified-site /note= "Potential phosphorylation site"  
 XX WO200034477-A2.  
 XX 15-JUN-2000.  
 XX 10-DEC-1999; 99WO-US30408.  
 XX 11-DEC-1998; 98US-0210083.  
 PR 11-DEC-1998; 98US-9123456.  
 PR 09-FEB-1999; 99US-0119365.  
 PR 16-MAR-1999; 99US-0124687.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;  
 PI Lu DAM, Azimzai Y;  
 XX WPI; 2000-423423/36.  
 DR N-PSDB; AAA47423.  
 XX New human neuron-associated proteins and polynucleotides encoding them,  
 PT useful for diagnosis, treatment and prevention of cell proliferative  
 PT disorders including cancer, neuronal and neurological disorders  
 XX Claim 1; Page 103-105; 145pp; English.  
 XX Human neuron-associated proteins (NEUAP) can be used for for  
 CC treating or preventing a disorder associated with decreased  
 CC expression or activity of NEUAP. Antagonists of NEUAP are useful for  
 CC treating or preventing disorder associated with increased expression  
 CC or activity of NEUAP. NEUAP or their fragments or derivatives are  
 CC useful for treating neurological disorder such as epilepsy, ischemic  
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's

CC disease, Pick's disease, Huntington's disease, dementia and  
 CC Parkinson's disease. NEUAPs are also useful for treating other  
 CC demyelinating diseases, bacterial and viral meningitis, prion  
 CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and  
 CC metabolic diseases of the nervous system, neurofibromatosis, other  
 CC developmental disorders of the central nervous system, cerebral  
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,  
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and  
 CC other neuromuscular disorders, peripheral nervous system disorders,  
 CC inherited, metabolic, endocrine, and toxic myopathies, mental  
 CC disorders including mood, anxiety and schizophrenic disorders, a cell  
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective  
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal  
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,  
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory  
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's  
 CC disease, adult respiratory distress syndrome, allergies, ankylosing  
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,  
 CC complications of cancer, hemodialysis, and extracorporeal circulation,  
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic  
 CC infections, and trauma. This protein was given the incyte ID no.  
 CC 2888437CDL.  
 XX  
 SQ Sequence 854 AA;

Query Match 1.4%; Score 19; DB 21; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 1e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 ALCADPKVKERRAHARQCL 962  
 Db 471 alcakdpvkerraharqcl 489  
 |||||

RESULT 5  
 AAB58462  
 ID AAB58462 standard; Protein; 363 AA.  
 AC AAB58462;  
 XX 14-MAR-2001 (first entry)  
 DT  
 XX Lung cancer associated polypeptide sequence SEQ ID 800.  
 DE  
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease.  
 XX  
 OS Homo sapiens.  
 XX WO200055180-A2.  
 PN  
 XX 21-SEP-2000.  
 PD  
 XX 08-MAR-2000; 2000WO-US05918.  
 PF  
 XX 12-MAR-1999; 99US-0124270.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C.A.  
 XX  
 XX Ruben SM;  
 PI  
 XX WPI; 2000-587514/55.  
 DR N-PSDB; AAF18338.  
 DR  
 XX Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -



Db 1 mknhsksgtstlrltlllthsdgdlteqgkiskpmsrlrlaagsaivklapepyhei 60

QY 899 ITLEQYQALAINDECYQVQVFAQLHRG 929

Db 61 itleqyqalaindecyqvrqvfqqlhkg 91

RESULT 3

AAV65247

ID AAY65247 standard; Protein; 101 AA.

XX AC AAY65247;

XX 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1408.

DE Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;

KW gene therapy; chromosome mapping; upstream regulatory sequence;

KW forensic; location; development; protein synthesis; stability;

KW regulation; identification.

XX Homo sapiens.

OS Homo sapiens.

PN W09953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-IB00712.

XX 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GIST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PI WPI; 2000-038446/03.

DR N-PSDB; AA242861.

XX Novel secreted protein 5' expressed sequence tag sequences used in

PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX Claim 3; Page 776; 837pp; English.

XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)

CC sequences, corresponding to human secreted proteins. AAY64651 to

CC AAY65438 represent the EST-related proteins corresponding to AA242265 to

CC AA243052. The 5' ESTs can be used for producing secreted human gene

CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the

CC location, development stage, rate, and quantity of protein synthesis, as

CC well as stability of mRNA. The ESTs are also useful as probes for

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can

CC also be used in forensic procedures to identify individuals, or in

CC diagnostic procedures to identify individuals having genetic diseases

CC resulting from abnormal gene expression. The products may also be used in

CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the

CC insertion of a polypeptide into a membrane, or importing a polypeptide

CC into a cell. The proteins encoded by the EST sequences may be useful in

CC treating a variety of human conditions. Secreted proteins have

CC therapeutic value, and the identification of new secreted proteins is

CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 101 AA;

Query Match 5.2%; Score 72; DB 21; Length 101;

Best Local Similarity 100.0%; Pred. No. 1.2e-67;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSKTRTNDGKITYPGPVKREISDKISKEMVRRLKNVVKTFMDMDQDSSEKELYNL 60

Db 1 mahsktrndgkitypgpvkreisdkiskeemvrllknvktfmdmdqdsseekelylnla 60

QY 61 LHLASDFELKHP 72

Db 61 lhlasdfelkhp 72

RESULT 4

AA01382

ID AAB01382 standard; Protein; 854 AA.

XX AC AAB01382;

XX 20-OCT-2000 (first entry)

XX Neuron-associated protein.

DE Neuron associated protein; NEUAP; neurological disorder; epilepsy;

KW ischemic cerebrovascular disease; stroke; cerebral neoplasm;

KW Alzheimer's disease; Pick's disease; Huntington's disease;

KW dementia; Parkinson's disease; demyelinating disease; meningitis;

KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;

KW cerebral palsy; muscular dystrophy; central nervous system; CNS;

KW peripheral nervous system; PNS; myopathy; schizophrenia;

KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;

KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;

KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;

KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;

KW AIDS; Addison's disease; adult respiratory distress syndrome;

KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;

XX Werner syndrome, trauma; human.

XX Homo sapiens.

XX OS

XX Key

PH Modified-site 48 Location/Qualifiers

FT /note= "Potential glycosylation site"

FT Modified-site 88

FT /note= "Potential phosphorylation site"

FT Modified-site 99

FT /note= "Potential phosphorylation site"

FT Modified-site 153

FT /note= "Potential phosphorylation site"

FT Modified-site 159

FT /note= "Potential glycosylation site"

FT Modified-site 175

FT /note= "Potential phosphorylation site"

FT Modified-site 249

FT /note= "Potential phosphorylation site"

FT Modified-site 265

FT /note= "Potential phosphorylation site"

FT Modified-site 330

FT /note= "Potential phosphorylation site"

FT Modified-site 340

FT /note= "Potential phosphorylation site"

FT Region 358..409

FT /label= "Postsynaptic protein signature"

FT Modified-site 369

FT /note= "Potential glycosylation site"

FT Region 373..422

FT /label= "Neuraxin signature"

FT Modified-site 375

FT /note= "Potential glycosylation site"

FT Modified-site 377

FT /note= "Potential phosphorylation site"

FT Modified-site 387

FT /note= "Potential phosphorylation site"

FT Modified-site 398

FT /note= "Potential phosphorylation site"

FT Modified-site 422

FT /note= "Potential phosphorylation site"



Db 121 kspqnrnyfyllenawksynaicfeledsneiftqlyrtlfsvinnghnqkvhnmhvd1 180  
QY 181 MSSITCEGDTVSQELLDVNLVPAHNLNKQAVDLAKALLKRTAQALPVTTFNQV 240  
Db 181 mssiccdgtvsqelltdvnlvpahnlknqaydlakallkrtaqalepyttcfinqv 240  
QY 241 LMLGKTSISDSERHFDLTILELYNDSDHLLSVLPQLEFKLSNDNEERLQVVKLARMF 300  
Db 241 lmlgktsisdsehrfdlilelyndshllsvlpqlfeklksndneerlqvkvllakmf 300  
QY 301 GAKDELSAQNKPLWCQYGRFNDIHVPIRECVKFAFASHCLMNHDPDLAKDITEYLKVRSH 360  
Db 301 gakselasnqplwqcygrfndihvpirlecvkfashclmnhdpdlakditeylkvrsh 360  
QY 361 DPEEARHDIIVSYTAAKDILLVNDHLLNFVRETDLDKRWVRKEAMGLAQIYKKA 420  
Db 361 dpeearhdiivsytaakdillvndhllnfvretdldkrwvrkeamglaqiyykka 420  
QY 421 LQSAAGKDAKQIAWKDKLLHIYQNSIDDRLLVERIFAQVMPVPHNLETTERMKCLYYL 480  
Db 421 lqsaagkdaakqiawkdkllhiyqnsidrrllverifagymvphnlettermkclyy1 480  
QY 481 YATDLNAVAKALNEMWKCONLRHQVDLIDLKOPKTDASVKAIFSKVMVITRNLPPDG 540  
Db 481 yatldnavakalnemwkconlrrhqvdildlikpdktdasvkaifskvmvitrnlppdg 540  
QY 541 KAQDPKKFTQVLEDEKIRKOLEVLVSPTCSCQAEGCVREITKKLGNPKQPTNPFLEM 600  
Db 541 kaqdpkkftqvledekirkolevlvsptcscqaegcvreitkklgnpkqptnplfem 600  
QY 601 TKFLERIAPVHIDPESALIKOVNKSIDGTADDEGVPTDQAIRAGLELLKVLSTFH 660  
Db 601 tkfleriapvhidpesalikvynksidgtaddegvptdqairaglellkvlstfh 660  
QY 661 PISFHSATFESLLACLKMDKVAEALQIFKNTGSKIEEDFPHIRSALLPVLHKKSK 720  
Db 661 pisfhsatfessllaclkmdkvaeealqifkntgskieedfphirsallpvlhkksk 720  
QY 721 GPPRAKYAIIHAIFFSKETQFAQIEPLHKSIDPSNLEHLITPLVTIGHIALAPDQ 780  
Db 721 gpprakyaiaihaiffsketqfaqieplhksidpsnlehlitplvtighiallapdq 780  
QY 781 FAAPKSWVATFIVKDLMLNDRLPGKTKTWLWVPDVEEYSPETMVKIQAKMVRWLLGMK 840  
Db 781 faapksvattfivkdlmlndrlpgkttklwvpdeevspetmviqakmvrwllgm 840  
QY 841 NNHKSSTGLRLTLILHSDGDLTEQKISKPDMSRLRLAAGSAIVKLAQEPVHEIIT 900  
Db 841 nnhksstglrltlilthsdgdlteqgiskpdmrslrlaagsaivklagepcyheii 900  
QY 901 LEQYQALCALINDECYQVROVFAQKLHKLRLPLEYMAICALCAKDPVKERRAHARQ 960  
Db 901 leqyqalcalindecyqvrovfaqklhklrlpleymaicalcakdpvkerraharq 960  
QY 961 CLVKNINVRREYLKQHAASEKLLSLLPEYVVPYTHLLAHDPDYVKQDTEQLKDVKEC 1020  
Db 961 clvkninvrreylkhaasekllsllpeyvvythllahdpdyvkqdlteqlkdvkec 1020  
QY 1021 LMFVLEILMAKNENSHAFIRKWNENIKQTKDAQGDDAKMNEKIYTVCDVAMNIIIMSKS 1080  
Db 1021 lmfvleilmaknenshafirkwnenikqtkdaqgddakmnekitytcdvamniiimsk 1080  
QY 1081 TTYTESPKDPVLPARFTQPKNSNTKNYLPPEMKSFETPKPKTKTNVLGAVNKPSS 1140  
Db 1081 ttytespkdpvlparfthqpknsntknylppemksfettppkpktktnvlgavnkplss 1140  
QY 1141 AGKQSTKSSRMETVSNASSNSPPGRIGKRLDSSEMDHSENEDEYTMSSPLPGKKSDDK 1200  
Db 1141 agkqstksrmetvsnassnsppgrigkrlsdssmdhseenedeytmssplpgkksdk 1200  
QY 1201 RDDSILVRSLEKPRGRKTPVTEQEEKLGMDDLTKLVOEQPKGQSQRKRGHTASBDS 1260  
Db 1201 rddsilvrselekprgrktpvteqeeeklgmdlltklvqeqpkgsqrsrkrhtasesd 1260

QY 1261 EQOWPEEKRLKEDILENEQNSPPKKGKGRPPKPLGGTPEKEPTMKTSKKSKKSKG 1320  
Db 1261 eqowpeekrlkedilenedeqnsppkkgkgrppkplggtpkeptmktkskksksg 1320  
QY 1321 PPAPEEEEEEROSGNTQKSKSKOHVRSRRAQRAESPESAIESTOSTPQKGRGRPSK 1380  
Db 1321 ppapeeeeeerqsgnteqkskskqhrrsrraqraespaesaiestqtpqkgrgrpsk 1380  
QY 1381 TPSPSQPKKNV 1391  
Db 1381 tpspsqpkknv 1391

## RESULT 2

AAG02811 standard; Protein; 92 AA.

XX AAG02811;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 6892.

XX Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC02817.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 6892; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
XX were prepared from total human RNAs or polyA+ RNAs derived from 30  
XX different tissues. EST sequences usually correspond mainly to the 3'  
XX untranslated region (UTR) of the mRNA because they are often obtained  
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
XX those cases where longer cDNA sequences have been obtained, the full 5'  
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
XX ends and can therefore be used to obtain full length cDNAs and genomic  
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
XX chromosome mapping procedures. They are used to obtain upstream  
XX regulatory sequences and to design expression and secretion vectors.

XX Sequence 92 AA;

Query Match

Best Local Similarity 6.5%; Score 91; DB 21; Length 92;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 MKNHKSCTSTLRLLTLILHSDGDLTEQKISKPDMSRLRLAAGSAIVKLAQEPYHEI 898

|||||



961 6 0.4 16 21 AAY67428  
962 6 0.4 16 21 AAY67429  
963 6 0.4 16 21 AAY58819  
964 6 0.4 16 21 AAY58820  
965 6 0.4 16 21 AAY58821  
966 6 0.4 17 15 AAR47023  
967 6 0.4 17 15 AAR63134  
968 6 0.4 17 17 AAW11068  
969 6 0.4 17 18 AAW39022  
970 6 0.4 17 18 AAW31116  
971 6 0.4 17 19 AAW47330  
972 6 0.4 17 20 AAY55857  
973 6 0.4 18 12 AAR15131  
974 6 0.4 18 13 AAR26378  
975 6 0.4 18 17 AAW11063  
976 6 0.4 18 18 AAW31115  
977 6 0.4 18 19 AAW47382  
978 6 0.4 18 20 AAY55869  
979 6 0.4 18 20 AAY42670  
980 6 0.4 18 20 AAY33094  
981 6 0.4 19 8 AAF70521  
982 6 0.4 19 18 AAW11745  
983 6 0.4 20 9 AAP81079  
984 6 0.4 20 21 AAY78405  
985 6 0.4 20 21 AAY58507  
986 6 0.4 20 22 AAB70459  
987 6 0.4 20 22 AAB55804  
988 6 0.4 20 22 AAB57666  
989 6 0.4 20 22 AAB58076  
990 6 0.4 21 18 AAW11765  
991 6 0.4 21 18 AAW11743  
992 6 0.4 22 20 AAY29727  
993 6 0.4 23 6 AAP50753  
994 6 0.4 24 21 AAB29673  
995 6 0.4 24 21 AAY52645  
996 6 0.4 25 13 AAR20169  
997 6 0.4 25 16 AAR95574  
998 6 0.4 25 21 AAR83358  
999 6 0.4 26 14 AAR33541  
1000 6 0.4 27 18 AAW06240

ALIGNMENTS

RESULT 1  
ID AAY94702 standard; Protein; 1391 AA.  
XX  
AC AAY94702;

01-DEC-2000 (first entry)

Human androgen shutoff gene 3 (AS3) protein sequence.

Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;  
chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;  
diagnosis; treatment; cytostatic; human; ss.

OS Homo sapiens.

Key Location/Qualifiers  
FH Region 419..459  
FT /label= "Subdomain I"  
FT /note= "Hank's conserved region"  
FT 419..425  
FT /label= "Beta strand 1"  
FT 426..433  
FT /label= "Mg-ATP binding loop"  
FT 453..459  
FT /label= "Beta strand 2"  
FT 472..482  
FT /label= "Subdomain II"

Histone H2B peptid  
Histone H2B peptid  
Aggrecan degrading  
Aggrecan degrading  
Aggrecan degrading  
K-ras transformati  
Egr-1 nuclear loca  
Atherosclerotic pl  
Peptide resembling  
Atherosclerotic pl  
Inhibitor of LDL b  
Apolipoprotein fra  
Vascular injury af  
Sequence of athero  
Atherosclerotic pl  
Atherosclerotic pl  
Inhibitor of LDL b  
Apolipoprotein fra  
HHV-6 variant B de  
Alipoprotein B der  
Antigenic oligopep  
T-cell receptor pe  
Sequence encoded b  
Human papillomavir  
HHV8 ORF K8.1-deri  
L. lactis signal p  
PL peptide #15. S  
CD105 PL peptide.  
CD105 PL peptide.  
T-cell receptor pe  
T-cell receptor pe  
Feline leukaemia v  
Sequence of a port  
Human/mouse SLE-as  
v-ras encoded onco  
Fusion junction of  
Protein Ki-ras 2A  
Conserved b region  
Mastoparan analog  
Variable heavy cha

FT Region  
FT /note= "Hank's conserved domain - beta strand 3"  
FT 489..498  
FT /label= "Subdomain IIR"  
FT /note= "Hank's conserved region - alpha helix C"  
FT 509..516  
FT /label= "Subdomain IV"  
FT /note= "Hank's conserved region - beta strand 4"  
FT 525..548  
FT /label= "Subdomain V"  
FT /note= "Hank's conserved region"  
FT 525..532  
FT /label= "Beta strand 5"  
FT 540..548  
FT /label= "Al[ha helix D"  
FT 554..567  
FT /label= "Subdomain Via"  
FT /note= "Hank's conserved region - alpha helix E"  
WO200050454-A1.  
31-AUG-2000.  
24-FEB-2000; 2000WO-US04732.  
24-FEB-1999; 99US-0121461.  
(TUFT ) TUFTS. COLLEGE.  
Soto AM, Sonnenschein C, Geck P, Szelel J;  
WPI; 2000-565451/52.  
N-PSDB; AAA28051, AAA28052.  
New human androgen-induced tumor suppressor cDNA sequence termed  
'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient  
diagnosis and treatment of prostate cancer -  
Claim 2; Fig 1; 152pp; English.

This invention relates to a human androgen-induced tumour suppressor cDNA  
sequence termed 'Androgen Shutoff Gene 3' (AS3). The AS3 gene is located  
on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell  
proliferation and use as a marker for the efficient diagnosis and  
treatment of prostate cancer. The invention includes AS3 cDNA and protein  
sequences, a vector comprising the cDNA sequence, a host cell transfected  
with the expression vector, and a method for producing an AS3 polypeptide  
comprising culturing the transfected cells. AS3 has cytostatic activity,  
and acts to suppress cell proliferation. The AS3 gene is useful as a  
marker for the efficient diagnosis and treatment of prostate cancer. The  
AS3 nucleic acid molecule can be used as a source of antisense agents for  
sequence specific modulation of gene expression. The AS3 protein may be  
used in the treatment of disorders caused by aberrant modification or  
mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene  
or aberrant post-translational modification of the AS3 protein. This  
sequence represents the human AS3 protein sequence.

XX Sequence 1391 AA;

Query Match 100.0%; Score 1391; DB 21; Length 1391;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSKTRTNDGKITPPGVKEISDKISKEMVRLKMWKVTMDMDQDSEEEKELYLNLA 60  
|||||  
Db 1 mahsktrtdgkitppgvkeiskiskemvrrllkmvktfmdmqdseeeekelylnla 60  
QY 61 LHLASDFFLKHGKOVRLLVACCLADIFRIYAPEAPYTPDKLKLDIFMFTROLKGLDT 120  
|||||  
Db 61 lhlasdfflkhpqkdvrlvaccladifriyapeapytpdklkldifmfttrlqkldt 120  
QY 121 KSPQFNRYFYLLENIAWVKSYNICFELSDNEIFTOLYTLFSVINNGHNQKVMHMDVL 180  
|||||











523	7	0.5	414	21	AAG19531	Arabidopsis thalia	596	7	0.5	520	18	AAW15265	Rat beta-amyloid p
524	7	0.5	415	14	AAR32922	AAMP-1. Homo sapi	597	7	0.5	520	21	AAG44273	Arabidopsis thalia
525	7	0.5	416	16	AAR66773	N-terminally trunc	598	7	0.5	523	21	AAG44273	Arabidopsis thalia
526	7	0.5	417	19	AAW47034	Human cartilage gp	599	7	0.5	525	20	AAV14540	Rat lipolysis stim
527	7	0.5	421	19	AAW53121	Amino acid sequenc	600	7	0.5	525	22	AAW59906	Rat leptin fragmen
528	7	0.5	421	19	AAW53122	Amino acid sequenc	601	7	0.5	526	20	AAV14546	Mouse lipolysis st
529	7	0.5	421	21	AAG05485	Arabidopsis thalia	602	7	0.5	526	22	AAW59909	Murine leptin frag
530	7	0.5	422	18	AAW11301	Perilipin B (NSEQO	603	7	0.5	528	19	AAW59882	Amino acid sequenc
531	7	0.5	422	21	AAW12394	Rat perilipin B pr	604	7	0.5	528	22	AAW61241	Human BFR protein.
532	7	0.5	422	21	AAW08255	Arabidopsis thalia	605	7	0.5	530	21	AAG22021	Arabidopsis thalia
533	7	0.5	422	21	AAW13172	Arabidopsis thalia	606	7	0.5	530	21	AAG50060	Arabidopsis thalia
534	7	0.5	423	21	AAG26392	Arabidopsis thalia	607	7	0.5	531	21	AAG50059	Arabidopsis thalia
535	7	0.5	424	21	AAG29585	Arabidopsis thalia	608	7	0.5	532	21	AAG05484	Arabidopsis thalia
536	7	0.5	425	21	AAG29584	Arabidopsis thalia	609	7	0.5	535	19	AAW61028	Murine guanine nuc
537	7	0.5	426	16	AAW67776	Cell adhesion prot	610	7	0.5	536	21	AAG05483	Arabidopsis thalia
538	7	0.5	429	21	AAG28833	Arabidopsis thalia	611	7	0.5	539	21	AAW87747	Human brain proteo
539	7	0.5	429	21	AAG35587	Arabidopsis thalia	612	7	0.5	542	18	AAW36490	Human TULP1 protei
540	7	0.5	430	21	AAG29583	Arabidopsis thalia	613	7	0.5	542	21	AAW26905	Human TULP1 protei
541	7	0.5	431	21	AAG43794	Arabidopsis thalia	614	7	0.5	542	22	AAW66342	Castor bean calnex
542	7	0.5	433	21	AAW39007	Human secreted pro	615	7	0.5	543	21	AAW41267	Arabidopsis thalia
543	7	0.5	433	21	AAG28832	Arabidopsis thalia	616	7	0.5	544	16	AAW82213	Talaromyces flavus
544	7	0.5	433	21	AAG53586	Arabidopsis thalia	617	7	0.5	544	19	AAW72757	Human doublelin. Ho
545	7	0.5	435	21	AAG39453	Arabidopsis thalia	618	7	0.5	544	20	AAW55960	Mouse STE20-relate
546	7	0.5	435	21	AAG40292	Arabidopsis thalia	619	7	0.5	544	21	AAW42894	Human ORFX ORF2658
547	7	0.5	438	22	AAB30849	Amino acid sequenc	620	7	0.5	544	21	AAW87746	Rat brain proteogl
548	7	0.5	439	21	AAW53797	Amino acid sequenc	621	7	0.5	547	19	AAW69240	Clone A073.3 prote
549	7	0.5	442	22	AAW20330	Human protein phos	622	7	0.5	549	21	AAG29189	Arabidopsis thalia
550	7	0.5	443	20	AAW92948	WO9905287 Seq ID 8	623	7	0.5	552	22	AAW75305	Human secreted pro
551	7	0.5	444	21	AAW42151	Human ORFX ORF1915	624	7	0.5	554	19	AAW61026	Murine guanine nuc
552	7	0.5	450	18	AAW29454	Streptococcus pneu	625	7	0.5	559	18	AAW43396	Human imidazoline
553	7	0.5	450	20	AAW68551	S. pneumoniae Murd	626	7	0.5	560	21	AAG29188	Arabidopsis thalia
554	7	0.5	452	21	AAW45369	Arabidopsis thalia	627	7	0.5	562	16	AAW70491	Leucocytozoan prot
555	7	0.5	453	21	AAW47481	Arabidopsis thalia	628	7	0.5	563	21	AAW16687	Bacteriophage Dp-1
556	7	0.5	453	21	AAW45368	Arabidopsis thalia	629	7	0.5	568	19	AAW51244	Human calcitonin r
557	7	0.5	456	19	AAW82841	Human cerebral pro	630	7	0.5	568	22	AAW31977	Amino acid sequenc
558	7	0.5	456	20	AAW35959	Extended human sec	631	7	0.5	570	18	AAW36002	Human Fchd531 gene
559	7	0.5	456	21	AAW19530	Arabidopsis thalia	632	7	0.5	570	21	AAW45013	Protein encoded by
560	7	0.5	463	21	AAW67500	PBR-associated pro	633	7	0.5	574	20	AAV14539	Rat lipolysis stim
561	7	0.5	465	15	AAW50951	Receptor for C-ter	634	7	0.5	574	22	AAW59905	Rat leptin fragmen
562	7	0.5	465	18	AAW15266	Beta-amyloid precu	635	7	0.5	575	20	AAV14545	Mouse lipolysis st
563	7	0.5	469	22	AAW59875	Carbon monoxide de	636	7	0.5	575	21	AAG36370	Arabidopsis thalia
564	7	0.5	471	21	AAW90287	Human peptidase, H	637	7	0.5	575	22	AAW59908	Murine leptin frag
565	7	0.5	477	15	AAW66067	Phosphorylated p66	638	7	0.5	580	21	AAG29848	Arabidopsis thalia
566	7	0.5	477	15	AAW43999	Human OTC. Homo s	639	7	0.5	593	20	AAV14538	Rat lipolysis stim
567	7	0.5	477	20	AAW27591	Human secreted pro	640	7	0.5	593	22	AAW59904	Rat leptin fragmen
568	7	0.5	479	20	AAW97108	Thermostable poly	641	7	0.5	594	14	AAW34936	CENP-B. Homo sapi
569	7	0.5	479	21	AAW29190	Arabidopsis thalia	642	7	0.5	594	20	AAV14544	Mouse lipolysis st
570	7	0.5	479	21	AAW32389	Herpesvirus entry	643	7	0.5	594	22	AAW59907	Murine leptin frag
571	7	0.5	479	22	AAW62022	P. furiosus recomb	644	7	0.5	598	20	AAW99065	Human U62317 prote
572	7	0.5	481	14	AAW33389	ATI/AVPv2 receptor	645	7	0.5	601	21	AAW79240	Human diarylsulfon
573	7	0.5	484	8	AAW71081	Sequence encoded b	646	7	0.5	605	21	AAW57950	Human transmembran
574	7	0.5	484	22	AAW30848	Amino acid sequenc	647	7	0.5	605	21	AAW79241	Human diarylsulfon
575	7	0.5	487	12	AAW14149	Human alpha 2 beta	648	7	0.5	612	21	AAG49327	Arabidopsis thalia
576	7	0.5	487	18	AAW11804	Human alpha-2b adr	649	7	0.5	618	18	AAW18571	Human cyclin D1-hu
577	7	0.5	489	21	AAG26391	Arabidopsis thalia	650	7	0.5	618	19	AAW47452	Human cyclin D1/cy
578	7	0.5	491	21	AAG22022	Arabidopsis thalia	651	7	0.5	618	21	AAG41266	Arabidopsis thalia
579	7	0.5	491	21	AAW50056	Arabidopsis thalia	652	7	0.5	619	13	AAW27651	Human calcium chan
580	7	0.5	491	21	AAW50051	Arabidopsis thalia	653	7	0.5	627	21	AAG36369	Arabidopsis thalia
581	7	0.5	493	21	AAW52778	Arabidopsis thalia	654	7	0.5	628	20	AAW37204	Amino acid sequenc
582	7	0.5	495	21	AAW80994	Soybean glycinin A	655	7	0.5	635	18	AAW19920	Human Ksr' (kinase
583	7	0.5	496	21	AAW41051	Human ORFX ORF815	656	7	0.5	635	19	AAW75855	Human secretory pr
584	7	0.5	502	21	AAW54580	A rat calcium/cal	657	7	0.5	635	20	AAW45160	Human secreted pro
585	7	0.5	503	19	AAW53232	Cryptosporidium pa	658	7	0.5	635	21	AAW71060	Human membrane tra
586	7	0.5	504	21	AAW52777	Arabidopsis thalia	659	7	0.5	637	18	AAW11704	High affinity Na+
587	7	0.5	505	19	AAW61027	Murine guanine nuc	660	7	0.5	637	19	AAW52294	Drosophila p70S6K
588	7	0.5	506	21	AAW52776	Arabidopsis thalia	661	7	0.5	638	17	AAW00092	CAMP phosphodiester
589	7	0.5	509	21	AAG08254	Arabidopsis thalia	662	7	0.5	642	20	AAW13396	Amino acid sequenc
590	7	0.5	510	20	AAW29618	Human PI3K protei	663	7	0.5	642	21	AAW33435	Human PRO332 prote
591	7	0.5	517	18	AAW11300	Perilipin A (NSEQO	664	7	0.5	642	22	AAW80264	Human PRO332 prote
592	7	0.5	517	21	AAW12393	Rat perilipin A pr	665	7	0.5	646	21	AAG39302	Arabidopsis thalia
593	7	0.5	517	21	AAW32390	Herpesvirus entry	666	7	0.5	647	18	AAW00926	Human cyclin D1-hu
594	7	0.5	519	21	AAW42468	Human ORFX ORF2232	667	7	0.5	647	19	AAW74575	Human cyclin D1/cy
595	7	0.5	519	22	AAW61330	Human transcriptio	668	7	0.5	647	21	AAW93992	Amino acid sequenc



377	7	0.5	296	21	AAG60360	Arabidopsis thalia	450	7	0.5	349	21	AA484617	Amino acid sequenc
378	7	0.5	297	21	AAG07037	Arabidopsis thalia	451	7	0.5	350	21	AAG31413	Arabidopsis thalia
379	7	0.5	298	21	AAG06458	Arabidopsis thalia	452	7	0.5	351	21	AAG29179	Arabidopsis thalia
380	7	0.5	299	21	AAG17712	Arabidopsis thalia	453	7	0.5	352	21	AAG18561	Arabidopsis thalia
381	7	0.5	299	21	AAG36935	Arabidopsis thalia	454	7	0.5	352	21	AAG49090	Arabidopsis thalia
382	7	0.5	300	21	AAG11577	Arabidopsis thalia	455	7	0.5	353	22	AAB30847	Amino acid sequenc
383	7	0.5	301	21	AAG19858	Arabidopsis thalia	456	7	0.5	356	18	AAW14284	Human neuro-D prot
384	7	0.5	302	21	AAG11576	Arabidopsis thalia	457	7	0.5	356	18	AAW22441	Human neurogenic d
385	7	0.5	302	21	AAG25491	Arabidopsis thalia	458	7	0.5	356	19	AAW71017	Human neuroDI prot
386	7	0.5	302	22	AAB88546	Haemophilus influe	459	7	0.5	356	19	AAW54042	Human BHF1 protein
387	7	0.5	303	21	AAG11575	Arabidopsis thalia	460	7	0.5	356	21	AB114351	Human Neurod 1 pro
388	7	0.5	303	21	AAG50935	Arabidopsis thalia	461	7	0.5	356	21	AB114346	Human Neurod1 prot
389	7	0.5	303	22	AAB46830	A. thaliana TTI pr	462	7	0.5	357	16	AAW77504	Murine neurogenic
390	7	0.5	304	13	AAW20062	Rat EDF-binding pr	463	7	0.5	357	18	AAW22436	Mouse neurogenic d
391	7	0.5	308	21	AAG19857	Arabidopsis thalia	464	7	0.5	357	19	AAW71009	Murine neuroDI pro
392	7	0.5	308	21	AAW73387	HTRM clone 3340290	465	7	0.5	357	21	AAG18560	Arabidopsis thalia
393	7	0.5	308	22	AAB60476	Human cell cycle a	466	7	0.5	357	21	AAG49089	Arabidopsis thalia
394	7	0.5	309	21	AAG25490	Arabidopsis thalia	467	7	0.5	358	21	AAG05935	Arabidopsis thalia
395	7	0.5	309	21	AAG29339	Arabidopsis thalia	468	7	0.5	358	21	AAG08294	Arabidopsis thalia
396	7	0.5	310	21	AAG19856	Arabidopsis thalia	469	7	0.5	360	17	AAW03627	Human follicle sti
397	7	0.5	310	21	AAG50934	Arabidopsis thalia	470	7	0.5	360	21	AAG37722	Arabidopsis thalia
398	7	0.5	312	21	AAG25489	Arabidopsis thalia	471	7	0.5	360	21	AAG45370	Arabidopsis thalia
399	7	0.5	313	20	AAW34857	Chlamydia pneumoni	472	7	0.5	361	21	AAG11457	Arabidopsis thalia
400	7	0.5	313	21	AAG50933	Arabidopsis thalia	473	7	0.5	361	21	AAG39266	Arabidopsis thalia
401	7	0.5	316	21	AAG47797	Arabidopsis thalia	474	7	0.5	363	19	AAW42427	PDC-E2, BCOADC-E2,
402	7	0.5	316	21	AAW77927	A. thaliana enviro	475	7	0.5	365	21	AAG37789	Arabidopsis thalia
403	7	0.5	317	11	AAW05163	Sequence of human	476	7	0.5	366	21	AAG23057	Arabidopsis thalia
404	7	0.5	317	19	AAW49038	Rabbit low density	477	7	0.5	369	21	AAG18559	Arabidopsis thalia
405	7	0.5	317	21	AB119760	Human bone staloer	478	7	0.5	369	21	AAG49088	Arabidopsis thalia
406	7	0.5	317	21	AAG47796	Arabidopsis thalia	479	7	0.5	371	21	AAG11456	Arabidopsis thalia
407	7	0.5	317	22	ABW87430	Human gene 11 enco	480	7	0.5	371	21	AAG39265	Arabidopsis thalia
408	7	0.5	321	21	AAG60359	Arabidopsis thalia	481	7	0.5	372	18	AAW25048	BRCA2 cancer suce
409	7	0.5	322	20	AAW85718	Novel protein (C10	482	7	0.5	373	21	AAW24242	Human Nogo B prote
410	7	0.5	323	21	AAG36934	Arabidopsis thalia	483	7	0.5	373	21	AAG36635	Arabidopsis thalia
411	7	0.5	323	21	AAG47795	Arabidopsis thalia	484	7	0.5	373	21	AAW56969	Human MAGI polypep
412	7	0.5	324	21	AAG06457	Arabidopsis thalia	485	7	0.5	373	21	AAW53624	A bone marrow secr
413	7	0.5	325	20	AAW35140	Chlamydia pneumoni	486	7	0.5	374	21	AAW58373	Lung cancer associ
414	7	0.5	327	21	AAG05936	Arabidopsis thalia	487	7	0.5	374	21	AAG34959	Arabidopsis thalia
415	7	0.5	328	20	AAW07097	Colon cancer assoc	488	7	0.5	379	21	AAG06328	Arabidopsis thalia
416	7	0.5	328	21	AAG30241	Arabidopsis thalia	489	7	0.5	380	21	AAW77814	Yeast Upf1 protein
417	7	0.5	328	21	AAG34960	Arabidopsis thalia	490	7	0.5	381	18	AAW22439	Human neurogenic d
418	7	0.5	329	20	AAW35146	Chlamydia pneumoni	491	7	0.5	381	19	AAW71015	Human neuroD2 prot
419	7	0.5	329	21	ABW42303	Human ORFX ORF2067	492	7	0.5	382	18	AAW22442	Mouse neurogenic d
420	7	0.5	329	21	AAG08296	Arabidopsis thalia	493	7	0.5	382	19	AAW71018	Murine neuroD2 pro
421	7	0.5	329	21	AAG18592	Arabidopsis thalia	494	7	0.5	382	21	ABW14347	CAMP phosphodie
422	7	0.5	329	21	AAG21956	Arabidopsis thalia	495	7	0.5	384	17	AAW00097	Human neurogenic b
423	7	0.5	329	21	AAG37723	Arabidopsis thalia	496	7	0.5	384	20	AAW49817	Human temporal lob
424	7	0.5	329	22	ABW36589	Human FLEXHT-11 pr	497	7	0.5	384	21	ABW20628	PGS18ARR human tem
425	7	0.5	330	21	ABW28231	RNA polymerase alp	498	7	0.5	386	17	AAW98521	Xenopus orphan rec
426	7	0.5	330	21	AAG29181	Arabidopsis thalia	499	7	0.5	386	21	AAG36634	Arabidopsis thalia
427	7	0.5	331	21	AAG07036	Arabidopsis thalia	500	7	0.5	386	21	AAG39454	Arabidopsis thalia
428	7	0.5	331	21	AAG21657	Arabidopsis thalia	501	7	0.5	386	21	AAG40293	Arabidopsis thalia
429	7	0.5	332	21	AAG36636	Arabidopsis thalia	502	7	0.5	388	21	AAG30240	Arabidopsis thalia
430	7	0.5	332	21	AAW81647	Streptococcus pneu	503	7	0.5	389	18	AAW35137	Arabidopsis thalia
431	7	0.5	333	21	AAG60983	Arabidopsis thalia	504	7	0.5	394	21	AAG08256	Arabidopsis thalia
432	7	0.5	334	19	AAW72171	HSV-2 strain SB5 C	505	7	0.5	398	21	ABW43929	Human cancer assoc
433	7	0.5	334	19	AAW72090	HSV-2 strain SB5 C	506	7	0.5	399	20	AAW28679	Human nh328_5 secr
434	7	0.5	334	19	AAW72005	HSV-2 strain SB5 C	507	7	0.5	399	22	ABW30846	Amino acid sequenc
435	7	0.5	335	19	AAW55117	Streptococcus pneu	508	7	0.5	400	21	AAG11455	Arabidopsis thalia
436	7	0.5	335	21	AAG60982	Arabidopsis thalia	509	7	0.5	400	21	AAG11455	Arabidopsis thalia
437	7	0.5	336	19	AAW98449	H. pylori GHP0 578	510	7	0.5	401	21	ABW58312	Lung cancer associ
438	7	0.5	336	19	AAW79094	Human secreted pro	511	7	0.5	402	21	AAG06327	Arabidopsis thalia
439	7	0.5	336	20	AAW49830	Protein sequence S	512	7	0.5	405	21	AAG13174	Arabidopsis thalia
440	7	0.5	337	20	AAW07754	Human secreted pro	513	7	0.5	408	21	AAW44211	Corn phytoene synt
441	7	0.5	339	21	ABW48519	Bacillus circulans	514	7	0.5	409	21	AAW47483	Arabidopsis thalia
442	7	0.5	339	21	AAG08295	Arabidopsis thalia	515	7	0.5	411	20	AAW68011	Yeast immunophilin
443	7	0.5	339	21	AAG29180	Arabidopsis thalia	516	7	0.5	411	21	AAW47482	Arabidopsis thalia
444	7	0.5	340	21	AAW54427	Arabidopsis thalia	517	7	0.5	412	17	AAW03626	Human thyrotropin
445	7	0.5	341	21	AAG21656	Arabidopsis thalia	518	7	0.5	412	20	AAW68010	S. frugiperda immu
446	7	0.5	343	18	AAW01561	MOC1 protein, Can	519	7	0.5	412	21	ABW09929	Human cancer regre
447	7	0.5	343	21	AAW15118	Candida cyclin dep	520	7	0.5	412	22	AAG19532	Arabidopsis thalia
448	7	0.5	345	21	ABW42984	Human ORFX ORF2748	521	7	0.5	412	22	ABW36604	Human FLEXHT-26 pr
449	7	0.5	349	19	AAW54365	Human retinitis p1	522	7	0.5	413	21	AAG13173	Arabidopsis thalia



231	7	0.5	200	21	AAG36740	Arabidopsis thalia	304	7	0.5	248	21	AAG57384	Arabidopsis thalia
232	7	0.5	200	22	AA876848	Human lung tumour	305	7	0.5	248	21	AA94901	Human secreted pro
233	7	0.5	200	22	AA876868	Human lung tumour	306	7	0.5	248	21	AA944629	Human myocardium s
234	7	0.5	201	19	AAW79540	Adenovirus serotyp	307	7	0.5	248	22	AA888487	Human membrane or
235	7	0.5	202	21	AAG30242	Arabidopsis thalia	308	7	0.5	248	22	AA888487	Human secretory pr
236	7	0.5	202	21	AA933263	A dopamine and cAM	309	7	0.5	249	18	AAW09404	Human hepatoma-der
237	7	0.5	204	21	AAG12954	Arabidopsis thalia	310	7	0.5	249	21	AAG29595	Arabidopsis thalia
238	7	0.5	204	21	AAG38260	Arabidopsis thalia	311	7	0.5	252	21	AA842156	Human ORFX ORF1920
239	7	0.5	205	21	AAG36739	Arabidopsis thalia	312	7	0.5	252	21	AAG18594	Arabidopsis thalia
240	7	0.5	206	21	AA999598	E. coli L-threonin	313	7	0.5	252	21	AAG21958	Arabidopsis thalia
241	7	0.5	207	21	AA843715	Human cancer assoc	314	7	0.5	254	21	AAG16726	Arabidopsis thalia
242	7	0.5	207	21	AA970950	Mouse oocyte-speci	315	7	0.5	254	21	AAG43924	Arabidopsis thalia
243	7	0.5	207	21	AA970951	Mouse ovary-specif	316	7	0.5	255	21	AA966717	Membrane-bound pro
244	7	0.5	207	22	AA863365	Human breast cance	317	7	0.5	255	22	AA865240	Human PRO1140 (UNO
245	7	0.5	209	21	AA990335	B. subtilis nitro	318	7	0.5	257	21	AAG31484	Arabidopsis thalia
246	7	0.5	209	21	AAG34568	Arabidopsis thalia	319	7	0.5	258	21	AAG12024	Arabidopsis thalia
247	7	0.5	210	21	AAG28047	Arabidopsis thalia	320	7	0.5	258	21	AAG54905	Arabidopsis thalia
248	7	0.5	211	20	AA925422	Human zcalc-1 prot	321	7	0.5	258	21	AA860845	Arabidopsis thalia
249	7	0.5	212	21	AAG07044	Arabidopsis thalia	322	7	0.5	259	22	AA87352	Human gene 11 enco
250	7	0.5	212	21	AAG42682	Arabidopsis thalia	323	7	0.5	261	21	AAG06329	Arabidopsis thalia
251	7	0.5	214	21	AA821047	Human nucleic acid	324	7	0.5	264	21	AA843838	Arabidopsis thalia
252	7	0.5	216	21	AA812136	Hydrophobic domain	325	7	0.5	265	19	AAW42381	Methanococcus igne
253	7	0.5	216	21	AA824036	Human PRO407 prot	326	7	0.5	265	21	AAG11735	Arabidopsis thalia
254	7	0.5	216	21	AAG11736	Arabidopsis thalia	327	7	0.5	265	21	AAG18593	Arabidopsis thalia
255	7	0.5	216	21	AAG43331	Arabidopsis thalia	328	7	0.5	265	21	AAG21957	Arabidopsis thalia
256	7	0.5	217	21	AAG17714	Arabidopsis thalia	329	7	0.5	265	21	AAG27773	Arabidopsis thalia
257	7	0.5	218	21	AAG04566	Arabidopsis thalia	330	7	0.5	265	21	AAG43330	Arabidopsis thalia
258	7	0.5	218	21	AA859303	Arabidopsis thalia	331	7	0.5	266	21	AAG16759	Arabidopsis thalia
259	7	0.5	219	21	AAG09137	Arabidopsis thalia	332	7	0.5	266	21	AAG23678	Arabidopsis thalia
260	7	0.5	220	21	AA853454	Human colon cancer	333	7	0.5	267	21	AAG36738	Arabidopsis thalia
261	7	0.5	220	21	AA856929	Human prostate can	334	7	0.5	268	20	AA938440	Human secreted pro
262	7	0.5	221	21	AAG07043	Arabidopsis thalia	335	7	0.5	268	20	AA900274	Human secreted pro
263	7	0.5	221	21	AA842681	Arabidopsis thalia	336	7	0.5	269	21	AA843837	Arabidopsis thalia
264	7	0.5	223	21	AA834567	Arabidopsis thalia	337	7	0.5	270	21	AAG27772	Arabidopsis thalia
265	7	0.5	223	21	AAG43796	Arabidopsis thalia	338	7	0.5	271	21	AAG16758	Arabidopsis thalia
266	7	0.5	223	21	AA92346	Human cancer assoc	339	7	0.5	271	21	AAG23677	Arabidopsis thalia
267	7	0.5	224	15	AA874004	Tumour rejection a	340	7	0.5	271	21	AAG43795	Arabidopsis thalia
268	7	0.5	224	16	AA882989	P815A antigen prec	341	7	0.5	273	19	AAW59646	Amino acid sequenc
269	7	0.5	224	17	AA894615	MAGE protein. Not	342	7	0.5	274	21	AAG31414	Arabidopsis thalia
270	7	0.5	224	20	AA922127	pLA protein sequen	343	7	0.5	274	21	AAG43836	Arabidopsis thalia
271	7	0.5	224	21	AA87165	Human secreted pro	344	7	0.5	275	21	AAG27771	Arabidopsis thalia
272	7	0.5	224	21	AA876198	Nuclear transport	345	7	0.5	276	21	AAG16757	Arabidopsis thalia
273	7	0.5	225	19	AA885950	S. pneumoniae deri	346	7	0.5	277	21	AAG50744	Arabidopsis thalia
274	7	0.5	225	21	AAG28046	Arabidopsis thalia	347	7	0.5	278	21	AAG14223	Arabidopsis thalia
275	7	0.5	226	21	AAG17113	Arabidopsis thalia	348	7	0.5	279	21	AAG12023	Arabidopsis thalia
276	7	0.5	226	21	AA848555	Arabidopsis thalia	349	7	0.5	279	21	AAG60644	Arabidopsis thalia
277	7	0.5	226	21	AA944630	Human mature myoca	350	7	0.5	281	21	AAG60643	Arabidopsis thalia
278	7	0.5	227	18	AAW36148	Rat eosinophil-der	351	7	0.5	281	22	AA848050	Signal transductio
279	7	0.5	227	20	AA925425	Human zcalc-1 prot	352	7	0.5	282	21	AAG05937	Arabidopsis thalia
280	7	0.5	227	21	AA858926	Breast and ovarian	353	7	0.5	282	21	AAG12022	Arabidopsis thalia
281	7	0.5	227	21	AA809136	Arabidopsis thalia	354	7	0.5	284	16	AA871461	Porcine surfactant
282	7	0.5	228	22	AA863361	Human breast cance	355	7	0.5	284	20	AA974157	Human prostate tum
283	7	0.5	229	21	AAG17250	Arabidopsis thalia	356	7	0.5	284	21	AAG37724	Arabidopsis thalia
284	7	0.5	231	21	AAG43925	zeae mays protein f	357	7	0.5	284	21	AA950330	Human clone vb22_1
285	7	0.5	231	22	AA820276	Mouse interleukin	358	7	0.5	285	20	AA942000	Mouse DNA demethyl
286	7	0.5	232	21	AA825144	Pinus radiata cell	359	7	0.5	286	21	AA841938	Human ORFX ORF1702
287	7	0.5	234	13	AA829717	Guinea pig MBP-2.	360	7	0.5	286	21	AAG13505	Arabidopsis thalia
288	7	0.5	234	19	AAW40076	Guinea pig eosinop	361	7	0.5	287	21	AAG13504	Arabidopsis thalia
289	7	0.5	234	21	AAG31415	Arabidopsis thalia	362	7	0.5	288	21	AA842132	Human ORFX ORF1896
290	7	0.5	234	22	AA820278	Mouse interleukin	363	7	0.5	291	20	AA932752	Short-chain tumour
291	7	0.5	236	15	AA856668	Bacteroides fragil	364	7	0.5	291	20	AA941498	Human DNA demethyl
292	7	0.5	240	20	AAW92295	Mouse alpha-1 (XVI	365	7	0.5	292	21	AAG12056	Arabidopsis thalia
293	7	0.5	240	21	AAG07042	Arabidopsis thalia	366	7	0.5	292	21	AAG42996	Arabidopsis thalia
294	7	0.5	240	21	AA842680	Arabidopsis thalia	367	7	0.5	293	20	AA931938	Mouse radi7 cell c
295	7	0.5	241	21	AAG29981	Arabidopsis thalia	368	7	0.5	293	21	AAG13503	Arabidopsis thalia
296	7	0.5	245	21	AAG38221	Arabidopsis thalia	369	7	0.5	294	21	AA860361	Arabidopsis thalia
297	7	0.5	246	21	AA857385	Arabidopsis thalia	370	7	0.5	295	13	AA827114	pradi. Homo sapie
298	7	0.5	247	21	AAG29980	Arabidopsis thalia	371	7	0.5	295	14	AA844802	Human cyclin D1
299	7	0.5	247	21	AA957952	Human transmembran	372	7	0.5	295	15	AA854044	bcl-1 protein. Ho
300	7	0.5	248	20	AA925421	Human zcalc-1 alle	373	7	0.5	295	21	AAG06459	Arabidopsis thalia
301	7	0.5	248	21	AA818922	A novel polypeptid	374	7	0.5	295	21	AAG20848	Arabidopsis thalia
302	7	0.5	248	21	AA824034	Human PRO4354 prot	375	7	0.5	295	21	AAG29340	Arabidopsis thalia
303	7	0.5	248	21	AAG10894	Arabidopsis thalia	376	7	0.5	295	21	AA851130	Arabidopsis thalia







GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.  
 OM protein - protein search, using sw model  
 Run on: September 25, 2001, 10:03:14 ; Search time 27.59 Seconds  
 (without alignments)  
 3056.469 Million cell updates/sec

Title: US-09-512-581-2  
 Perfect score: 1391  
 Sequence: 1 MAHSKTRTDGKITPPGVK.....QKGRGPKSPTSPSPKKNV 1391

Scoring table: OLIGO  
 Gapop 60.0 , Gapext 60.0  
 Searched: 412676 seqs, 60623988 residues  
 Word size : 0

Total number of hits satisfying chosen parameters: 412676  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_0601.\*  
 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1391	100.0	1391	21	AA194702 Human androgen shu
2	91	6.5	92	21	AA194702 Human secreted pro
3	72	5.2	101	21	AA195247 Human 5' EST relat
4	19	1.4	854	21	AA191382 Neuron-associated
5	14	1.0	363	21	AA195846 Lung cancer associ
6	9	0.6	67	20	AA1902188 Polypeptide sequen
7	9	0.6	67	22	AA1966299 Human TRF1 TANK1 b
8	9	0.6	134	21	AA1940441 Human secreted pro
9	9	0.6	325	19	AA197351 Altered telomere r
10	9	0.6	437	19	AA197350 Altered telomere r
11	9	0.6	439	18	AA1928555 Telomeric repeat b

12	9	0.6	439	19	AA197348 Human telomere rep
13	9	0.6	439	20	AA1902182 A human telomere r
14	9	0.6	439	21	AA1903942 Human telomeric re
15	9	0.6	439	22	AA1966298 Human TRF1 SEQ ID
16	9	0.6	1005	21	AA1901833 Haemophilus influe
17	9	0.6	1011	21	AA1901832 Haemophilus influe
18	9	0.6	1886	19	AA1954241 Rattus norvegicus
19	9	0.6	26	14	AA1912333 Prepro-thyrotropin
20	8	0.6	26	19	AA1973133 Corticotropin rele
21	8	0.6	26	19	AA1981900 Human CRF protein
22	8	0.6	26	22	AA1961357 Human corticotropi
23	8	0.6	28	14	AA1912334 Prepro-thyrotropin
24	8	0.6	32	17	AA195846 Human prepro-TRH p
25	8	0.6	58	21	AA196078 Arabidopsis thalia
26	8	0.6	58	21	AA1960357 Arabidopsis thalia
27	8	0.6	74	21	AA1941122 Zea mays protein f
28	8	0.6	81	21	AA1919544 Arabidopsis thalia
29	8	0.6	81	21	AA1937862 Arabidopsis thalia
30	8	0.6	97	21	AA1939617 Arabidopsis thalia
31	8	0.6	149	21	AA1919543 Arabidopsis thalia
32	8	0.6	149	21	AA1937861 Arabidopsis thalia
33	8	0.6	158	21	AA1931670 Arabidopsis thalia
34	8	0.6	159	21	AA1937860 Arabidopsis thalia
35	8	0.6	165	21	AA1939616 Arabidopsis thalia
36	8	0.6	232	13	AA1929716 Guinea pig MBP-1.
37	8	0.6	233	19	AA1940075 Guinea pig eosinop
38	8	0.6	237	21	AA1916483 Arabidopsis thalia
39	8	0.6	237	21	AA1937419 Arabidopsis thalia
40	8	0.6	242	14	AA1931229 Prepro-thyrotropin
41	8	0.6	242	20	AA1968538 Prepro-TRH protein,
42	8	0.6	242	22	AA1920275 Human interleukin
43	8	0.6	243	22	AA1920277 Human interleukin
44	8	0.6	254	21	AA1916482 Arabidopsis thalia
45	8	0.6	254	21	AA1937418 Arabidopsis thalia
46	8	0.6	260	21	AA1908527 Arabidopsis thalia
47	8	0.6	260	21	AA1938888 Arabidopsis thalia
48	8	0.6	294	21	AA1931669 Arabidopsis thalia
49	8	0.6	307	21	AA1916481 Arabidopsis thalia
50	8	0.6	307	21	AA1937417 Arabidopsis thalia
51	8	0.6	313	21	AA1908526 Arabidopsis thalia
52	8	0.6	313	21	AA1938887 Arabidopsis thalia
53	8	0.6	338	21	AA1908525 Arabidopsis thalia
54	8	0.6	338	21	AA1938886 Arabidopsis thalia
55	8	0.6	346	21	AA1947606 Arabidopsis thalia
56	8	0.6	351	21	AA1918312 Plasmodium falci
57	8	0.6	451	13	AA1929636 PCTD ORF 1. Chlam
58	8	0.6	468	21	AA1952008 P. horikoshii PHB
59	8	0.6	468	21	AA1951637 P. horikoshii PHB
60	8	0.6	470	21	AA193902 Mouse HS6ST2 prote
61	8	0.6	473	21	AA1956316 Human secreted pro
62	8	0.6	505	20	AA1949903 Mouse ROSA26 antis
63	8	0.6	531	20	AA1942226 Human pancreatic d
64	8	0.6	568	20	AA1935121 Amino acid sequenc
65	8	0.6	605	20	AA1931741 Human podocalyxin-
66	8	0.6	648	21	AA1950048 Candida albicans p
67	8	0.6	653	17	AA1989903 Murine APLP1. Mus
68	8	0.6	677	20	AA1928815 PL776_6 secreted p
69	8	0.6	709	19	AA1953345 Human adenylyl cycl
70	8	0.6	710	20	AA1980995 Human guanine nucl
71	8	0.6	823	21	AA190295 Human peptidase, H
72	8	0.6	993	20	AA1949897 Rat TAO2 kinase.
73	8	0.6	1062	21	AA1940294 Human ORFX ORF58 p
74	8	0.6	1165	14	AA1937309 Cardiac adenylyl c
75	8	0.6	1168	20	AA1930599 Human type VI aden
76	8	0.6	1180	21	AA1902010 Type VI adenylyl c
77	8	0.6	1235	21	AA1941663 Human ORFX ORF1427
78	8	0.6	1857	21	AA1953970 Human periphera b
79	7	0.5	9	22	AA1970243 Polytopic peptide
80	7	0.5	10	22	AA1970238 Polytopic peptide
81	7	0.5	12	16	AA192654 IML-1 alpha-E chal
82	7	0.5	12	20	AA1955318 Glutamate rich pep
83	7	0.5	12	22	AA1946774 Peptide E4E fragme
84	7	0.5	15	19	AA1951573 Delta-sleep induci



**THIS PAGE BLANK (USPTO)**







Search completed: September 25, 2001, 22:57:52  
Job time: 23890 sec



Human; P2; CX5C chemokine; Chromosome 5q31; gene therapy; asthma; allergic rhinitis; urticaria; anaphylactic shock; hives; hay fever; ds.







and cells the genes are expressed in. Examples of activities include: antiarthritic; immunosuppressive; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neurotrophic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; and vulnerary. The polynucleotides and polypeptides can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders, cardiovascular disorders, cerebrovascular disorders, angiogenesis, nervous system disorders, infections caused by bacteria, viruses and fungi and ocular disorders. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. AAC79731 to AAC79739 and AAB44594 represent sequences used in the exemplification of the present invention.

Sequence 794 BP; 236 A; 182 C; 168 G; 208 T; 0 other;

Query Match 0.5%; Score 26; DB 21; Length 794;

Best Local Similarity 100.0%; Pred. No. 19; Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4436 tttttttttttttttttttttttgtc 4461

|||||

Db 785 TTTT

RESULT 44

AAT27152/c

ID AAT27152 standard; DNA; 807 BP.

XX

AC

XX

XX

DT

DB

XX

XX

KW

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

XX

PR

XX

XX

PA

XX

XX

PI

XX

XX

DR

XX

XX

PT

XX

XX

Human Machado-Joseph disease-related gene fragment 2.  
Human: Machado-Joseph disease; mature protein; repeat motif; probe; cerebral temporal fossa lobe cortex, ss.  
Synthetic.  
JP08092289-A.  
09-APR-1996.  
21-SEP-1994; 94JP-0251600.  
21-SEP-1994; 94JP-0251600.  
(ONOY ) ONO PHARM CO LTD.  
Kakizuka A;  
WPI; 1996-236099/24.  
Human Machado-Joseph disease-related protein and DNA encoding it -  
used in the diagnosis and treatment of MJD  
Claim 8; Page 10; 12pp; Japanese.  
This is the nucleotide sequence of a human Machado-Joseph disease-related gene fragment. The complete gene (AAT27151) was isolated from a human cerebral temporal fossa lobe cortex mRNA-derived cDNA library using the probe AAT27160. 8 highly positive clones were isolated including the clone CAG-27 (AAT26151). The BamHI-BglII and DraII-SacI

fragments of this clone were then used to isolate other clones (sequences not in specification). The clones or fragments can be used in the diagnosis and treatment of Machado-Joseph disease.

Sequence 807 BP; 271 A; 171 C; 170 G; 195 T; 0 other;

Query Match 0.5%; Score 26; DB 17; Length 807;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4435 cttttttttttttttttttttttgtt 4460

|||||

Db 270 CTTT

RESULT 45

AAZ52545/c

ID AAZ52545 standard; cDNA; 1067 BP.

XX

AC

XX

DT

XX

XX

DE

XX

XX

KW

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

XX

XX

Human secreted protein clone yk14\_1 nucleotide sequence SEQ ID NO:141.  
Human: secreted protein; immunostimulatory; haemostatic; cytokine; proliferative; differentiative; chemotactic; chemokinetic; vaccine; thrombolytic; antiinflammatory; cytostatic; immunosuppressive; gene therapy; ss.

Sequence 807 BP; 271 A; 171 C; 170 G; 195 T; 0 other;

Query Match 0.5%; Score 26; DB 17; Length 807;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4435 cttttttttttttttttttttttgtt 4460

|||||

Db 270 CTTT

RESULT 45

AAZ52545/c

ID AAZ52545 standard; cDNA; 1067 BP.

XX

AC

XX

DT

XX

XX

DE

XX

XX

KW

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

XX

XX

Human secreted protein clone yk14\_1 nucleotide sequence SEQ ID NO:141.  
Human: secreted protein; immunostimulatory; haemostatic; cytokine; proliferative; differentiative; chemotactic; chemokinetic; vaccine; thrombolytic; antiinflammatory; cytostatic; immunosuppressive; gene therapy; ss.

Sequence 807 BP; 271 A; 171 C; 170 G; 195 T; 0 other;

Query Match 0.5%; Score 26; DB 17; Length 807;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4435 cttttttttttttttttttttttgtt 4460

|||||

Db 270 CTTT

RESULT 45

AAZ52545/c

ID AAZ52545 standard; cDNA; 1067 BP.

XX

AC

XX

DT

XX

XX

DE

XX

XX

KW

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

PR

XX

XX

XX

XX

Human secreted protein clone yk14\_1 nucleotide sequence SEQ ID NO:141.  
Human: secreted protein; immunostimulatory; haemostatic; cytokine; proliferative; differentiative; chemotactic; chemokinetic; vaccine; thrombolytic; antiinflammatory; cytostatic; immunosuppressive; gene therapy; ss.

Sequence 807 BP; 271 A; 171 C; 170 G; 195 T; 0 other;

Query Match 0.5%; Score 26; DB 17; Length 807;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4435 cttttttttttttttttttttttgtt 4460

|||||

Db 270 CTTT

RESULT 45

AAZ52545/c

ID AAZ52545 standard; cDNA; 1067 BP.

XX

AC

XX

DT

XX

XX

DE

XX

XX

KW

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

PF







KW Colon cancer; detect; differential expression; human; treatment;  
 KW detect mutation; non-invasive diagnostic method; ds.  
 OS Homo sapiens.  
 XX WO200012702-A2.  
 XX 09-MAR-2000.  
 XX 30-AUG-1999; 99WO-US19424.  
 XX 31-AUG-1998; 98US-0098639.  
 PR 27-JAN-1999; 99US-0117393.  
 XX (FARB ) BAYER CORP.  
 XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;  
 PI Catino TU, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;  
 PI Schlegel R;  
 XX WPI; 2000-256641/22.  
 DR Novel nucleic acids and proteins for identifying therapeutic agents  
 PT useful for treating and diagnosing cancer, especially colon cancer -  
 XX  
 XX Claim 16; Page 280; 345pp; English.  
 XX This sequence represents a human nucleotide sequence which is  
 CC differentially expressed in colon cancer cells compared to the expression  
 CC levels in normal cells. The nucleotide sequence can be used as a source  
 CC of primers and probes. The nucleotide sequence is useful for determining  
 CC the phenotype of a cell by detecting the differential expression of the  
 CC sequence relative to a normal cell. The probes derived from the sequence  
 CC can also be used to determine the phenotype of cells in a sample. Probes  
 CC and antibodies which hybridise to the nucleotide sequence can also be  
 CC used to determine the phenotype of a cell. The primers are useful for  
 CC detecting a mutation in a test nucleotide sequence and also for detecting  
 CC cancer, preferably colon cancer. Antibodies against the protein encoded  
 CC by the nucleotide sequence can also be used in a method to detect colon  
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
 CC colon cancer at an early stage.  
 XX  
 SQ Sequence 535 BP; 136 A; 124 C; 127 G; 139 T; 9 other;

Query Match 0.5%; Score 26; DB 21; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4435 cttttttttttttttttttttttttgt 4460  
 Db 5 cttttttttttttttttttttttttgt 30

RESULT 40  
 AAQ75445/C  
 ID AAQ75445 standard; DNA; 635 BP.  
 XX  
 AC AAQ75445;

XX 07-AUG-1995 (first entry)

XX Murine glycosylation inhibiting factor (GIF).

XX Glycosylation inhibiting factor; GIF; ss.

XX Mus musculus.

XX Key Location/Qualifiers  
 XX CDS 82..426  
 XX /\*tag= a

XX WO9426923-A.

XX 24-NOV-1994.  
 XX 13-MAY-1994; 94WO-US05354.  
 XX 14-MAY-1993; 93US-0061041.  
 XX (KIRI ) KIRIN BEER KK.  
 PA (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.  
 XX Ishizaka K, Liu Y, Mikayama T;  
 PI WPI; 1995-006814/01.  
 DR P-PSDB; AAR63858.  
 XX New fusion polypeptide(s) and glycosylation inhibition factor -  
 PT used to develop prods. for suppressing an undesirable response,  
 PT partic. allergies and auto-immune disease  
 XX Claim 38; Figure 1; 156 pp; English.  
 XX Murine GIF was purified from culture supernatant of GIF-producing  
 CC T cell hybridoma 231F1 cells using anti-lipomodulin monoclonal  
 CC antibody 141B9. The N-terminal AA sequence was directly sequenced  
 CC (see AAR63857). Based on the N-terminal AA sequence and the sequence  
 CC of another peptide (AN-5 AAR63854), oligos were synthesised.  
 CC Attempts were made to amplify a partial cDNA by PCR, and to use  
 CC the cDNA to probe a murine cDNA library. The synthesised primers  
 CC used in the PCR were AAQ75443 and AAQ75444. The PCR amplified a 0.2 Kb  
 CC fragment which was used to screen murine T cell hybridoma, 231F1  
 CC cDNA library. The longest clone (0.65 Kb) was chosen for DNA  
 CC sequencing. The nucleotide sequence and deduced AA sequence of  
 CC murine GIF are shown in AAQ75445/63858. Estimated size of GIF  
 CC protein is 13 kDa which correlates with that of purified GIF from  
 CC the T hybridoma 231F1 cells. Murine GIF protein lacks a signal  
 CC peptide.  
 XX  
 SQ Sequence 635 BP; 183 A; 192 C; 142 G; 118 T; 0 other;

Query Match 0.5%; Score 26; DB 16; Length 635;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4436 tttttttttttttttttttttttttg 4461  
 Db 577 tttttttttttttttttttttttttg 552

RESULT 41  
 AAT38390/C  
 ID AAT38390 standard; cDNA; 635 BP.  
 XX  
 AC AAT38390;

XX 02-JAN-1997 (first entry)

XX Murine glycosylation inhibiting factor cDNA.

XX Antigen-specific glycosylation inhibiting factor; AgGIF; lymphokine;  
 KW T-cell receptor; TCR; immunosuppressant; immunosuppressive;  
 KW hypersensitivity; graft rejection; autoimmune disease; ss.

XX Mus sp.

XX Key Location/Qualifiers  
 XX CDS 82..429  
 XX /\*tag= a  
 XX /product= full-length murine GIF

XX WO9631617-A1.

XX 10-OCT-1996.



human colon cancer differentially expressed microRNA sequence #393:



XX	
Dt	14-MAY-1998 (first entry)
XX	
DE	3' portion of cDNA clone encoding secreted protein AS32.
XX	
KW	Human; secreted protein; research; treatment; AS32;
KW	3' portion; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO9739122-A2.
XX	
PJ	23-OCT-1997.
XX	
PA	11-APR-1997; 97WO-US06042.
XX	
PR	12-APR-1996; 96US-0631184.
XX	
PPA	(MURO-) MURO PHARM INC.
XX	
PT	Theoharides TC;
XX	
DR	WPI; 1997-526459/48.
XX	
PT	Human and murine secreted proteins - useful to research or treat
XX	diseases or disorders related to their function
PS	Disclosure; Page 99; 140pp; English.
CC	
CC	The present sequence is the 3' portion of a cDNA clone encoding a
CC	human secreted protein, which may have nutritional uses, or
CC	cytokine and cell proliferation/differentiation, immune stimulating
CC	or suppressing, haematopoiesis regulating, tissue growth,
CC	activin/inhibin, chemotactic/chemokinetic, haemostatic and
CC	thrombolytic, receptor/ligand, anti-inflammatory or tumour
CC	inhibition activities. It can also be used to research or treat
CC	diseases/disorders related to its function.
CC	The partial cDNA clone AP162 was 1st isolated from a human adult
CC	placenta cDNA library. The partial cDNA clones AM931, AM610, AM340,
CC	AM282, AK647, AK583, AK533 and AK296 were 1st isolated from a human
CC	fetal kidney cDNA library. The partial cDNA clones H617 and BB9
CC	were 1st isolated from a human peripheral blood monocyte cell (Th1
CC	or Th2) cDNA library. The partial cDNA clone AW191 was 1st isolated
CC	from a human ovary (PA-1 teratocarcinoma) cDNA library. The partial
CC	cDNA clones AT211, AT205 and AT319 were 1st isolated from a human
CC	lymphocyte and dendritic cell cDNA library. The partial cDNA clones
CC	AS34 and AS32 were 1st isolated from a human foetal brain cDNA
CC	library. The partial cDNA clone AR260 was 1st isolated from a human
CC	adult retina cDNA library. The partial cDNA clones K640 and K39
CC	were 1st isolated from a murine bone marrow (stromal cell line
CC	FCM-4) cDNA library.
XX	
SQ	Sequence 394 BP; 111 A; 95 C; 113 G; 64 T; 11 other;
	Query Match 0.5%; Score 26; DB 18; Length 394;
	Best Local Similarity 100.0%; Pred. No. 22;
	Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	4434 tctttttttttttttttttttg 4459
Db	375 TC TTTTTTTTTTTTTTTTTTGGG 350
	RESULT 36
	AAV02161/c
ID	RAV02161 standard; cDNA; 394 BP.
XX	
AQ	AAV02161;
XX	
DT	12-MAY-1998 (first entry)
XX	
DE	Human secreted protein AS32 3' portion including the poiva tail.

XX	Human; secreted protein; ATCC 98026; cytokine; immunomodulation;
KW	cell proliferation; differentiation; regulation; ds.
OS	Homo sapiens.
XX	
XX	WO9739123-A2.
PN	
PD	23-OCT-1997.
XX	
XX	14-APR-1997; 97WO-US06139.
PF	
XX	
PR	18-APR-1996; 96US-0634325.
XX	
PA	(GEMY ) GENETICS INST INC.
PI	Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA;
PI	Spaulding V;
XX	
DR	WPI; 1997-526460/48.
XX	
PT	New secreted proteins encoded clones present in ATCC 98026 -
PT	possibly having cytokine, cell proliferation/differentiation
PT	. regulating, immunomodulating and many other activities
XX	
PS	Disclosure; Page 99; 139pp; English.
XX	
CC	The present sequence encodes a portion of a novel human secreted protein
CC	deposited under accession number ATCC 98026. The secreted protein can be
CC	used to determine biological activity, to raise antibodies, as tissue
CC	markers, to isolate cognate ligands or receptors, to identify agents
CC	that modulate their interactions and as nutritional supplements. It may
CC	also have a very wide range of biological activities although no
CC	evidence for any is provided in the specification. Typical of these are
CC	cytokine, cell proliferation/differentiation modulating activity or
CC	induction of other cytokines; immunostimulating/immunosuppressant
CC	activities (e.g. for treating human immunodeficiency virus infection,
CC	cancer, autoimmune diseases and allergy); regulation of haematopoiesis
CC	(e.g. for treating anaemia or as adjunct to chemotherapy); stimulation
CC	of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for
CC	treating wounds, periodontal disease, neurological diseases stroke,
CC	fibrosis); inhibition or stimulation of follicle stimulating hormone
CC	(for control of fertility); chemotactic and chemokinetic activities
CC	(e.g. for treating infections, tumours); haemostatic or thrombolytic
CC	activity (e.g. for treating haemophilia, cardiac infarction etc.);
CC	anti-inflammatory activity (e.g. for treating septic shock, Crohn's
CC	disease); as antimicrobials; for treating psoriasis or other
CC	hyperproliferative diseases; for regulation of metabolism, behaviour, and
CC	many others. Also contemplated is the use of the corresponding nucleic
CC	acid in gene therapy procedures.
XX	
SQ	Sequence 394 BP; 111 A; 95 C; 113 G; 64 T; 11 other;
Query Match 0.5%; Score 26; DB 18; Length 394;	
Best Local Similarity 100.0%; Pred. No. 22;	
Matches	26; Conservative 0; Mismatches 0; Indels 0; Gaps
OY	4434 tccttttttttttttttttttttgg 4459 
Db	375 TCCTTTTTCCTTTTTCCTTTTTCCTTTGG 350
RESULT	37
AAC93441/c	ID
AAC93441 standard; cdNA; 451 BP.	ID
XX	AAC93441;
XX	
DT	16-FEB-2001 (first entry)
XX	
DE	Human secreted protein gene 20 SEQ ID NO:30.
XX	

XX		Human; secreted protein; ATCC 98026; cytokine; immunomodulation;
KW		cell proliferation; differentiation; regulation; ds.
OS		Homo sapiens.
XX		
XX		
PN		WO9739123-A2.
PD		23-OCT-1997.
XX		
XX		14-APR-1997; 97WO-US06139.
PF		
XX		
PR		18-APR-1996; 96US-0634325.
XX		(GEMY ) GENETICS INST INC.
PA		
PI	Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA;	
PI	Spaulding V;	
XX		WPI; 1997-526460/48.
DR		
XX		New secreted proteins encoded clones present in ATCC 98026 -
PT		possibly having cytokine, cell proliferation/differentiation
PT		. regulating, immunomodulating and many other activities
XX		
XX		Disclosure; Page 99; 139pp; English.
PS		The present sequence encodes a portion of a novel human secreted protein
CC		deposited under accession number ATCC 98026. The secreted protein can be
CC		used to determine biological activity, to raise antibodies, as tissue
CC		markers, to isolate cognate ligands or receptors, to identify agents
CC		that modulate their interactions and as nutritional supplements. It may
CC		also have a very wide range of biological activities although no
CC		evidence for any is provided in the specification. Typical of these are
CC		cytokine, cell proliferation/differentiation modulating activity or
CC		induction of other cytokines; immunostimulating/immunosuppressant
CC		activities (e.g. for treating human immunodeficiency virus infection,
CC		cancer, autoimmune diseases and allergy); regulation of haematopoiesis
CC		(e.g. for treating anaemia or as adjunct to chemotherapy); stimulation
CC		of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for
CC		treating wounds, periodontal disease, neurological diseases stroke,
CC		fibrosis); inhibition or stimulation of follicle stimulating hormone
CC		(for control of fertility); chemotactic and chemokinetic activities
CC		(e.g. for treating infections, tumours); haemostatic or thrombolytic
CC		activity (e.g. for treating haemophilia, cardiac infarction etc.);
CC		anti-inflammatory activity (e.g. for treating septic shock, Crohn's
CC		disease); as antimicrobials; for treating psoriasis or other
CC		hyperproliferative diseases; for regulation of metabolism, behaviour, and
CC		many others. Also contemplated is the use of the corresponding nucleic
CC		acid in gene therapy procedures.
XX		
SQ		Sequence 394 BP; 111 A; 95 C; 113 G; 64 T; 11 other;
Query Match 0.5%; Score 26; DB 18; Length 394;		
Best Local Similarity 100.0%; Pred. No. 22;		
Matches	26; Conservative 0; Mismatches 0; Indels 0; Gaps	
OY	4434 tcttttttttttttttttttttttggg 4459	
Db	375 TCCTTTTTCCTTTTTTTTTTTGGTGG 350	
RESULT	37	
AAC93441/c		
ID	AAC93441 standard; cDNA; 451 BP.	
XX	AAC93441;	
XX		
DT	16-FEB-2001 (first entry)	
DE	Human secreted protein gene 20 SEQ ID NO:30.	
XX		



Df  
76 acgttttcttttttttttttttg 101

RESULT 34  
AAV89475/c  
ID AAV89475 standard; DNA; 293 BP.  
XX AC AAV89475;  
XX DT 13-APR-1995 (first entry)  
XX XX Junction A of SfiI fragment contg. Newcastle Disease Virus genes.  
XX DB SfiI fragment; Newcastle Disease Virus; junction; ss.  
XX KW Synthetic.  
OS Key Location/Qualifiers  
FH misc\_feature 1..39 /tag= a  
FT label= linker  
FT promoter 40..78 /tag= b  
FT EP1  
FT EP1  
FT promoter 79..121 /tag= c  
FT LP2  
FT misc\_feature 122..182 /tag= d  
FT NDV  
PN W09419014-A.

XX X  
X PD 01-SRP-1994.  
PF PF 28-FEB-1994; 94WO-USO1826.  
PR PR 26-FEB-1993; 93US-0024156.  
PA PA (JAFG ) NIPPON ZEON KK.  
PI PI (SYTR ) SYNTRO CORP.  
PS Cochran MD;

XX WIPI: 1994-294007/36.  
XX New recombinant fowl pox virus for use in vaccines - contains  
PT PT genes expressing antigens of Newcastle disease virus and opt.  
PT PT infectious bronchitis virus  
PP Disclosure; Figure 1; 85pp; English.

PQ Newcastle Disease Virus (NDV) HN and F genes were inserted as a  
CC SfiI fragment into the homology vector 443-88.8 at the unique  
CC SfiI site. The NDV HN and F genes were inserted in the same  
CC transcriptional orientation as the ORF in the parental homologous  
CC vector. The sequence of SfiI fragment is in AAQ68943/R58858/AAQ58859.  
CC The inserted SfiI fragment has the following structure:  
CC Junction A - Fragment 1 (HN), AAs 2-577) - Junction B - Fragment  
CC 2 (F, AAs 1-553) - Junction C - Fragment 3 (pBR322) - Junction D.  
CC Fragment 1 is approx. 1811 bp Avail to NaeI fragment of the NDV HN  
CC cDNA clone (B1 strain). Fragment 2 is an approx 1812 bp BamHI to  
CC PstI fragment of the full length NDV F cDNA (B1 Strain). Fragment  
CC 3 is an approx 235 bp PstI and SacI fragment of the plasmid pBR322.  
CC The sequences of the Junctions are in AAQ68945, AAQ68946, AAQ68947 and  
CC AAQ68948.

XX Sequence 182 BP; 52 A; 33 C; 42 G; 55 T; 0 other:



```

FT primer_bind /*tag= as
FT complement (6772..6790)
FT /*tag= at
FT /note= "downstream amplification primer 10-504"
FT 7120..7137
FT /*tag= au
FT /note= "upstream amplification primer 10-204"
FT 7422..7468
FT /*tag= av
FT /note= "potential binding site for a probe"
FT replace (7445, A)
FT /*tag= aw
FT /note= "claim 4"
FT 7513..7531
FT /*tag= ax
FT /note= "upstream amplification primer 10-32"
FT complement (7557..7574)
FT /*tag= ay
FT /note= "downstream amplification primer 10-204"
FT 7612..7637
FT /*tag= az
FT /note= "specifically claimed in claim 3"
FT 7709..7852
FT /*tag= ba
FT /number= 1
FT 7783..78290
FT /*tag= bb
FT /product= FLAP
FT /note= "contains introns"
FT 7847..7893
FT /*tag= bc
FT /note= "potential binding site for a probe"
FT 7853..78235
FT /*tag= bd
FT /number= 1
FT replace (7870, A)
FT /*tag= be
FT /note= "claim 4"
FT complement (7914..7933)
FT /*tag= bf
FT /note= "downstream amplification primer 10-32"
FT 8117..715994
FT /*tag= bg
FT /note= "specifically claimed in claim 1"
FT 16114..16132
FT /*tag= bh
FT /note= "upstream amplification primer 10-33"
FT 16236..16335
FT /*tag= bi
FT /number= 2
FT 16265..16311
FT /*tag= bj
FT /note= "potential binding site for a probe"
FT replace (16288, T)
FT /*tag= bk
FT /note= "claim 4"
FT 16324..16370
FT /*tag= bl
FT /note= "potential binding site for a probe"
FT 16336..24226
FT /*tag= bm
FT /number= 2
FT replace (16347, A)
FT /*tag= bn
FT 16360..16406
FT primer_bind

```

```

Query Match 0.5%; Score 27; DB 21; Length 43069;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 4432 aatcttttttttttttttttttttttgg 4458
Db 29038 AATCTTTTTTTTTTTTTTTTTTTTGTG 29002

```

## RESULT 32

AAD00579

ID AAD00579 standard; DNA; 128 BP.

XX AAD00579;

XX AAD00579;

XX 21-SEP-2000 (first entry)

XX Human Hscdc6 intron 2 of genomic sequence.

XX Hscdc6; CSH; human; DNA replication; mitosis; cyclin-dependent kinase;  
 KW cdk; regulator; inhibitor; cdc6p; cdc18; vertebrate cdc6; antagonist;  
 KW agonist; treatment; cell proliferative disease; psoriasis; medicament;  
 KW atherosclerotic vascular disease; vascular restenosis; cardiomyopathy;  
 KW inflammatory arthritis; autoimmune disease; organ transplant rejection;  
 KW traumatic injury; stroke; myocardial infarction; renal failure; antibody;  
 KW hepatic failure; cancer; breast; colon; cervix; lymphoma; diagnosis;  
 KW screen; cytostatic; immunosuppressive; antiinflammatory; antiarthritic;  
 KW antipsoriatic; antiarteriosclerosis; vasotropic; cardiac; vulnery;  
 KW tranquiliser; cerebroprotective; intron; ds.

XX Homo sapiens.

XX WO2000026242-A2.

XX 11-MAY-2000.

XX 28-OCT-1999; 99WO-US25445.

XX 30-OCT-1998; 98US-0183266.

XX (COLD-) COLD SPRING HARBOR LAB.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Stillman B, Williams RS, Mendez JR;

XX WPI; 2000-365574/31.

XX DNA replication-regulating gene, which is vertebrate cdc6 gene and its  
 PT antagonist is useful for treating cell proliferative diseases such as  
 PT atherosclerotic vascular disease, vascular restenosis, psoriasis

PS Claim 5; Fig 8B; 11pp; English.

XX The present DNA sequence is the intron 2 of human Hscdc6 genomic sequence  
 CC (previously referred as CSH), involved in cell cycle regulation. Hscdc6  
 CC functions as a regulator of DNA replication and/or entry of a cell into  
 CC mitosis and as a cyclin-dependent kinase (cdk) inhibitor. It is expressed  
 CC in active proliferative cells. It has similarity to cdc18 and cdc6p  
 CC proteins from Saccharomyces cerevisiae and S. pombe, respectively. The  
 CC agonist or antagonist of vertebrate cdc6 gene is useful for treating cell  
 CC proliferative diseases like atherosclerotic vascular disease, vascular  
 CC restenosis, psoriasis, inflammatory arthritis, autoimmune diseases and  
 CC organ transplant rejection. Cdc6 gene or agonist is useful for treating  
 CC conditions associated with loss of viable tissues, such as traumatic  
 CC injury, stroke, myocardial infarction, cardiomyopathy, renal and hepatic  
 CC failure, by enhancing cell proliferation. An antagonist comprising cdc6  
 CC specific antibody is useful to manufacture medicament for treatment of  
 CC cancer of the breast, colon, cervix or lymphoma. The antibodies are used  
 CC to screen for hscdc6 and diagnosis of a proliferative disorder.

XX Sequence 128 BP; 24 A; 22 C; 19 G; 63 T; 0 other;

```

Query Match 0.5%; Score 26; DB 21; Length 128;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 4434 tcttttttttttttttttttttttttgg 4459
Db 103 tcttttttttttttttttttttttttgg 128

```



Query Match 0.5%; Score 27; DB 22; Length 28690;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4433 atctttttttttttttttttttttttttttttgg 4459  
 |||||  
 Db 22470 atctttttttttttttttttttttttttttgg 22496

RESULT 31  
 AA236335/C  
 ID AA236335 standard; DNA; 43069 BP.  
 XX AA236335;  
 XX  
 XX 22-FEB-2000 (first entry)  
 XX  
 XX Genomic sequence of the 5-lipoxygenase activating protein (FLAP).  
 DE Human; 5-lipoxygenase activating protein; FLAP; biallelic marker;  
 XX leukotriene pathway; genotype; haplotype; FLAP-related biallelic marker;  
 KW asthma; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..7007  
 FT /tag= a  
 FT /note= "specifically claimed in claim 1"  
 FT misc\_feature 1..7708  
 FT /tag= b  
 FT /note= "potential 5' regulatory region"  
 FT primer\_bind 3851..3869  
 FT /tag= c  
 FT /note= "upstream amplification primer 10-517"  
 FT allele replace (3950, C)  
 FT primer\_bind 3927..3973  
 FT /tag= d  
 FT /tag= e  
 FT /note= "potential binding site for a probe"  
 FT primer\_bind complement (4171..4189)  
 FT /tag= f  
 FT /note= "downstream amplification primer 10-517"  
 FT primer\_bind 4120..4138  
 FT /tag= g  
 FT /note= "upstream amplification primer 10-518"  
 FT primer\_bind 4220..4266  
 FT /tag= h  
 FT /note= "potential binding site for a probe"  
 FT allele replace (4243, T)  
 FT /tag= i  
 FT primer\_bind 4289..4335  
 FT /tag= j  
 FT /note= "potential binding site for a probe"  
 FT allele replace (4312, G)  
 FT /tag= k  
 FT primer\_bind complement (4372..4390)  
 FT /tag= l  
 FT /note= "downstream amplification primer 10-518"  
 FT primer\_bind 4373..4391  
 FT /tag= m  
 FT /note= "upstream amplification primer 10-253"  
 FT primer\_bind 4467..4513  
 FT /tag= n  
 FT /note= "potential binding site for a probe"  
 FT allele replace (4490, G)  
 FT /tag= o  
 FT primer\_bind 4647..4693  
 FT /tag= p  
 FT /note= "potential binding site for a probe"  
 FT primer\_bind 4664..4710  
 FT /tag= q  
 FT /note= "potential binding site for a probe"

FT allele replace (4670, G)  
 FT /tag= r  
 FT allele replace (4687, C)  
 FT /tag= s  
 FT primer\_bind complement (4773..4792)  
 FT /tag= t  
 FT /note= "downstream amplification primer 10-253"  
 FT primer\_bind 4814..4833  
 FT /tag= u  
 FT /note= "upstream amplification primer 10-499"  
 FT primer\_bind 4945..4991  
 FT /tag= v  
 FT /note= "potential binding site for a probe"  
 FT primer\_bind 4956..4972  
 FT /tag= w  
 FT /note= "upstream amplification primer 10-500"  
 FT allele replace (4968, A)  
 FT /tag= x  
 FT primer\_bind complement (5026..5043)  
 FT /tag= y  
 FT /note= "downstream amplification primer 10-499"  
 FT primer\_bind 5117..5163  
 FT /tag= z  
 FT /note= "potential binding site for a probe"  
 FT allele replace (5140, T)  
 FT /tag= aa  
 FT primer\_bind 5190..5236  
 FT /tag= ab  
 FT /note= "potential binding site for a probe"  
 FT allele replace (5213, G)  
 FT /tag= ac  
 FT primer\_bind 5341..5387  
 FT /tag= ad  
 FT /note= "potential binding site for a probe"  
 FT allele replace (5364, A)  
 FT /tag= ae  
 FT primer\_bind complement (5405..5422)  
 FT /tag= af  
 FT /note= "downstream amplification primer 10-500"  
 FT primer\_bind 5524..5542  
 FT /tag= ag  
 FT /note= "upstream amplification primer 10-522"  
 FT primer\_bind 5571..5617  
 FT /tag= ah  
 FT /note= "potential binding site for a probe"  
 FT allele replace (5594, A)  
 FT /tag= ai  
 FT primer\_bind complement (5978..5996)  
 FT /tag= aj  
 FT /note= "downstream amplification primer 10-522"  
 FT primer\_bind 6218..6235  
 FT /tag= ak  
 FT /note= "upstream amplification primer 10-503"  
 FT primer\_bind 6347..6393  
 FT /tag= al  
 FT /note= "potential binding site for a probe"  
 FT allele replace (6370, G)  
 FT /tag= am  
 FT primer\_bind 6522..6539  
 FT /tag= an  
 FT /note= "upstream amplification primer 10-504"  
 FT primer\_bind complement (6652..6672)  
 FT /tag= ao  
 FT /note= "downstream amplification primer 10-503"  
 FT primer\_bind 6670..6716  
 FT /tag= ap  
 FT /note= "potential binding site for a probe"  
 FT allele replace (6693, T)  
 FT /tag= aq  
 FT primer\_bind 6740..6786  
 FT /tag= ar  
 FT /note= "potential binding site for a probe"  
 FT allele replace (6763, A)



```
PS Disclosure; Fig 1; 89pp; English.
```

```
xx CC The Glucanase II gene encodes the endo-beta-1,4-glucanase II enzyme.
```

```
CC The protein can be expressed in filamentous fungi, plant and yeast
```

```
CC systems. The enzyme can be used to degrade glucans. This can be used
```

```
CC for the preparation of foodstuffs, including feeds, as well as in the
```

```
CC brewing, pulp and textile industries and in the formulation of
```

```
CC detergents.
```

```
xx CC Sequence 3447 BP; 852 A; 733 C; 944 G; 918 T; 0 other;
```

```
SQ
```

```
Query Match 0.5%; Score 27; DB 19; Length 3447;
```

```
Best Local Similarity 100.0%; Pred. No. 6.9;
```

```
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4434 tcttttttttttttttttttgggt 4460
```

```
Db 197 tcttttttttttttttttttgggt 223
```

```
RESULT 30
```

```
AAF57118
```

xx  
xx  
AC  
xx  
xx  
DT

AA Human IL4Ralpha gene.  
DE  
XX  
KW Polymorphism: human: interleukin 4 receptor-alpha: IL4R-alpha:

XX	Homo sapiens.	
OS		
XX		
XX	WO200104270-A1.	
PN		
XX		
PD	18-JAN-2001.	
XX		
XX	13-JUL-2000; 2000WO-US19094.	
PF		
XX		
PR	13-JUL-1999; 99US-0143435.	
XX		
PA	(GENA-) GENAISSANCE PHARM INC.	
XX		
PI	Chew A., Denton RR., Duda A., Nandabalan K., Stephens JC.	

PI	Windemuth AK;
XX	
XX	WPI: 2001-103078/11.
DR	P-PSDB; AAB66970.
DR	
XX	
XX	New isolated polynucleotide useful for the identification of
PT	therapeutics in allergic diseases is new -
PT	
XX	
XX	Claim 1; Page 94-104; 188pp; English.
PS	
XX	
XX	The present invention relates to polymorphisms of the human interleukin 4
CC	receptor-alpha gene (IL4R-alpha). The present sequence is the reference
CC	sequence for the human IL4R-alpha gene. Polynucleotides comprising
CC	polymorphic gene variants are useful for therapeutic purposes. For
CC	example, where a patient may benefit from expression of a particular
CC	IL4Ralpha protein isoform, an expression vector encoding the isoform may
CC	be administered to the patient. It may desirable to decrease or block
CC	expression of a particular IL4Ralpha isogene, which may be done by
CC	turning off by transforming a targeted organ, tissue or cell population
CC	with an expression vector that expresses high levels of untranslatable
CC	mRNA for the isogene. Specific therapeutics identified by these methods
CC	

CC may be useful for allergic diseases.  
 XX  
 SQ sequence 28690 BP; 7026 A; 6911 C; 7548 G; 7205 T; 0 other;















```

PR 18-DEC-1997; 97US-0068054.
PR 18-DEC-1997; 97US-0068057.
PR 18-DEC-1997; 97US-0068064.
PR 18-DEC-1997; 97US-0070923.
PR 19-DEC-1997; 97US-0068169.
PR 19-DEC-1997; 97US-0068365.
PR 19-DEC-1997; 97US-0068367.
PR 19-DEC-1997; 97US-0068368.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;
PI Florence K, Greene JM, Janat F, Kyaw H, Moore PA;
PI Ni J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;
PI Yu G;
XX
DR WPI; 1999-418749/35.
DR P-PSDB; AAY36310.
XX
PT New isolated human genes encoding secreted polypeptides
PT Claim 1; Page 324; 537pp; English.
XX
CC AAX97916 to AAX98029 represent 110 isolated human secreted protein
CC genes. AAY36224 to AAY36727 represent the secreted proteins encoded by
CC the 110 human genes. The genes and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new genes.
CC Specific uses are described for each of the 110 genes, based on which
CC tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours, developmental
CC abnormalities and foetal deficiencies, blood disorders, diseases of the
CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
CC disorders, kidney disorders, digestive/endocrine disorders, infections
CC and AIDS. The polypeptides are also useful for identifying their binding
CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
CC used in the exemplification of the present invention.
XX
SQ Sequence 1352 BP; 474 A; 259 C; 248 G; 371 T; 0 other;

Query Match 0.5%; Score 27; DB 20; Length 1352;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 4431 aaatctttttttttttttttttt 4457
Db 317 aaatctttttttttttttttttt 343

RESULT 23
AAC59286
ID AAC59286 standard; cDNA; 1358 BP.
XX
AC AAC59286;
XX
XT 02-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA #10.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulncary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WQ2000056753-Al.

```



DR WPI; 2000-594630/56.  
 XX P-PSDB; AAB39350.  
 XX  
 PT New nucleic acid molecules encoding 48 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 PS Claim 1; Page 343; 395pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the  
 CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to  
 CC AAB39400 represent human secreted polypeptide sequences and proteins  
 CC homologous to them, which are given in the exemplification of the present  
 CC invention. Human secreted proteins have activities based on the tissues  
 CC and cells the genes are expressed in. Examples of activities include:  
 CC antithrombotic; immunosuppressive; antirheumatic; antiproliferative;  
 CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;  
 CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also be used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
 CC disorders. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
 CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in  
 CC the exemplification of the present invention.  
 XX  
 SQ Sequence 680 BP; 208 A; 128 C; 109 G; 235 T; 0 other;

Query Match 0.5%; Score 27; DB 21; Length 680;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4434 tctttttttttttttttttttttgt 4460  
 Db | ttttttttttttttttttttttttttttttt  
 653 TCCTTTTCTTTTCTTTTCTTTTGT 627  
 RESULT 20  
 AAX40124  
 ID AAX40124 standard; DNA; 704 BP.  
 XX  
 AC AAX40124;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Gastric cancer associated gene.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9904265-A2.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98WO-US14679.  
 XX  
 XX 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR

PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PA Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 XX WPI; 1999-132448/11.  
 DR  
 XX New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67; Page 705; 787pp; English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 704 BP; 142 A; 144 C; 197 G; 202 T; 19 other;  
 Query Match 0.5%; Score 27; DB 20; Length 704;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4434 tctttttttttttttttttttttgt 4460  
 Db | ttttttttttttttttttttttttttttttt  
 19 tctttttttttttttttttttttgt 45  
 RESULT 21  
 AAC59318/c  
 ID AAC59318 standard; cDNA; 845 BP.  
 XX  
 AC AAC59318;  
 XX  
 DT 02-FEB-2001 (first entry)  
 XX  
 DE Human secreted protein cDNA #42.  
 XX  
 KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200056753-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 16-MAR-2000; 2000WO-US06765.  
 XX  
 XX 23-MAR-1999; 99US-0126051.  
 PR 10-DEC-1999; 99US-0169906.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA







Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;  
chronic ear disease; autoimmune disease; ss.

Homo sapiens.

WO200050593-A1.

31-AUG-2000.

25-FEB-2000; 2000WO-US04795.

25-FEB-1999; 99US-0121549.

(UYTE-) UNIV TENNESSEE RES CORP.

Yoo TJ;

WPI; 2000-558400/51.

New beta-tubulin antigen in the membranous structure of the inner ear,  
reactive with antibodies of patients with Meniere's disease, for  
diagnosing Meniere's disease and distinguishing this disease from other  
autoimmune ear diseases

Claim 3; Page 97-103; 115pp; English.

The present sequence encodes a beta-tubulin antigen. The protein is  
an antigen of the membranous structure of the inner ear protein, and  
is reactive with antibodies from patients having Meniere's disease.  
Meniere's disease is a chronic ear disease with unknown etiology.  
Serum from patients suffering from this disease contain autoantibodies  
against a 30 kDa cochlear protein antigen. The disease is believed to be  
an autoimmune disease. The beta-tubulin antigen is useful as a target  
substance in diagnosing or detecting Meniere's disease and in  
distinguishing this disease from other autoimmune ear diseases.

Sequence 14784 BP; 5454 A; 2966 C; 2926 G; 3438 T; 0 other;

Query Match 0.6%; Score 29; DB 21; Length 14784;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4430 aaatctttttttttttttttttttttttttttttg 4458

|||||

Db 5592 AAAATCTTTTTTTTTTTTTTTTTTTTGTG 5564

RESULT 16

AZ46508/c

ID AZ46508 standard; DNA; 45265 BP.

XX AZ46508;

13-MAR-2000 (first entry)

Sequence of a COSMID (GenBank #AF002998).

3'-5' exonuclease; gene therapy; protocols. antineoplastic; antiviral;  
chemotherapeutic; cell killing; human; TREX1; EXO 1; COSMID; ss.

Homo sapiens.

WO9961064-A1.

02-DEC-1999.

14-MAY-1999; 99WO-US10578.

22-MAY-1998; 98US-0083617.

12-NOV-1998; 98US-0191470.

(UYWA-) UNIV WAKE FOREST.

XX

PI

XX

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Perrino FW;

WPI; 2000-097077/08.

Novel genes used in hybridization assays to detect the capacity of  
cells to express exonucleases

Example 5; Page 56-83; 93pp; English.

The invention relates to human and mouse genes encoding 3'-5'  
exonucleases. The exonuclease proteins of the invention can be used to  
identify inhibitors and effectors of exonuclease activity. Specific  
binding polypeptides, e.g. antibodies, can be used for purifying  
exonuclease products and detection and quantification of exonuclease  
products in fluid and tissue samples using immunological procedures.

Binding proteins are also useful in modulating the activity of  
exonucleases. Polynucleotides of the invention are useful in  
hybridization assays to detect the capacity of cells to express  
exonucleases. They are also useful as the basis for diagnostic methods  
useful for identifying a genetic alteration in an exonuclease locus that  
underlies a disease state. Nucleic acids that modulate the expression of  
the exonuclease genes, e.g. antisense nucleic acids, ribozymes, triple  
helix oligonucleotide, can be used in gene therapy protocols. Resistance  
or ineffectiveness of certain antineoplastic and antiviral agents may be  
due to an exonuclease activity. A need exists for the identification of  
metabolic factors which modulate the ability of chemotherapeutic agents  
to effect cell killing. This need is met by the present invention. The  
exonuclease polypeptides of the invention can be used to design and  
identify therapeutics which increase efficiency of the chemotherapeutic  
agent at lower doses, which are more easily tolerated in patients, and  
reducing side effects. The present sequence represents a COSMID  
(GenBank #AF002998) comprising a 45kb human genomic sequence.

Sequence 45265 BP; 10218 A; 13218 C; 12660 G; 9169 T; 0 other;

Query Match 0.6%; Score 29; DB 21; Length 45265;

Best Local Similarity 100.0%; Pred. No. 0.95;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4433 atctttttttttttttttttttttttttttg 4461

|||||

Db 26287 ATCTTTTTTTTTTTTTTTTTTTTGTG 26259

RESULT 17

AAC23080

ID AAC23080 standard; cDNA; 60 BP.

XX AAC23080;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 27155.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST ) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

XX



```

XX SQ Sequence 530 BP; 191 A; 90 C; 83 G; 166 T; 0 other;

Query Match 1.8%; Score 97; DB 21; Length 530;
Best Local Similarity 100.0%; Pred. No. 8.3e-21;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 agatggttgtaaaacttttatgatatggaccaggactctgaagaagaaaggagcttt 231
   |||||||
DB 215 AGATGGTGTGAACACTTTATGATATGGACAGGACTCTGAAGAAGAAAGAGCTTT 156
   |||||||

QY 232 atttaaacctagctttacatcttgcctcagattttt 268
   |||||||
DB 155 ATTTAAACCTAGCTTACATCTGCTTCAGATTTTT 119

RESULT 13
AAAX21198
ID AAX21198 standard; DNA; 421 BP.
XX
AC AAX21198;
XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN WO9859034-A2.
XX
PD 30-DEC-1998.
XX
PF 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
PT WPI; 1999-081273/07.
XX
DR New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
PS Claim 1; Page 1119; 1150pp; English.
XX
CC AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 421 BP; 96 A; 86 C; 100 G; 139 T; 0 other;

Query Match 0.6%; Score 29; DB 20; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4429 aaaaactcttttttttttttttttttttttttttttt 4457
   |||||||
DB 188 aaaaactcttttttttttttttttttttttttttttt 216

RESULT 14
AAV87322
ID AAV87322 standard; CDNA; 489 BP.
XX
AC AAV87322;
XX
DT 27-APR-1999 (first entry)
XX
DE EST clone BP810.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845435-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GEM) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070076/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
PS Claim 1; Page 527-528; 633pp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity, haemostatic
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
XX
SQ Sequence 489 BP; 159 A; 95 C; 116 G; 119 T; 0 other;

Query Match 0.6%; Score 29; DB 20; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4429 aaaaactcttttttttttttttttttttttttttttt 4457
   |||||||
DB 139 aaaaactcttttttttttttttttttttttttttttt 167

RESULT 15
AAA64141/c
ID AAA64141 standard; DNA; 14784 BP.
XX
AC AAA64141;
XX
DT 20-DEC-2000 (first entry)
XX
DE Nucleotide sequence of a beta-tubulin antigen.
XX

```



SQ Sequence 150 BP; 43 A; 26 C; 38 G; 42 T; 1 other;  
Query Match 1.9%; Score 99; DB 19; Length 150;  
Best Local Similarity 99.3%; Pred. NO. 2.8e-21;  
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4491 cacaaatgggactgctgaagagtgacaggttgacatttacttggtagcccccatacatt 4550  
Db 1 cacaaatgggactgctgaagagtgacaggttgacatttacttggtagcccccatacatt 60  
QY 4551 ttgtggtcacatctttagccatacacatgtaacattgacttgaggtcttcttgaaagt 4610  
Db 61 ttgtggtcacatctttagccatacacatgtaacattgacttgaggtcttcttgaaagt 120  
QY 4611 taatgtcgatggctatgtagacataaaga 4640  
Db 121 taatgtcgatggctatgtagacataaaga 150  
RESULT 11  
AAZ12050  
ID AAZ12050 standard; DNA; 150 BP.  
XX AC AAZ12050;  
XX DT 30-MAR-1999 (first entry)  
XX DE Human biallelic polymorphic DNA fragment WI-18387a.  
XX KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
XX OS Homo sapiens.  
XX PN WO9820165-A2.  
XX PD 14-MAY-1998.  
XX PF 05-NOV-1997; 97WO-US20313.  
XX PR 06-NOV-1996; 96US-0030455.  
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX PI Hudson T, Lander ES, Wang D;  
XX DR WPI; 1998-286974/25.  
XX PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
XX PS Claim 1; Page 216; 310pp; English.  
XX CC AAX10269-X12937 are human DNA fragments which contain biallelic  
CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
CC can be used in methods for determining polymorphic forms in an individual  
CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular

CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
XX SQ Sequence 150 BP; 43 A; 26 C; 38 G; 42 T; 1 other;  
Query Match 1.9%; Score 99; DB 19; Length 150;  
Best Local Similarity 99.3%; Pred. NO. 2.8e-21;  
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4491 cacaaatgggactgctgaagagtgacaggttgacatttacttggtagcccccatacatt 4550  
Db 1 cacaaatgggactgctgaagagtgacaggttgacatttacttggtagcccccatacatt 60  
QY 4551 ttgtggtcacatctttagccatacacatgtaacattgacttgaggtcttcttgaaagt 4610  
Db 61 ttgtggtcacatctttagccatacacatgtaacattgacttgaggtcttcttgaaagt 120  
QY 4611 taatgtcgatggctatgtagacataaaga 4640  
Db 121 taatgtcgatggctatgtagacataaaga 150  
RESULT 12  
AAZ80598/c  
ID AAZ80598 standard; cDNA; 530 BP.  
XX AC AAZ80598;  
XX DT 07-APR-2000 (first entry)  
XX DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:682.  
XX KW Human; gene expression product; diagnosis; tumour; colon cancer;  
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
KW hyperplasia; ds.  
XX OS Homo sapiens.  
XX PN WO9964576-A2.  
XX PD 16-DEC-1999.  
XX PF 09-JUN-1999; 99WO-IB01062.  
XX PR 10-JUN-1998; 98US-0088801.  
XX PA (FARB ) BAYER CORP.  
XX PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
PI Schlegel R;  
XX DR WPI; 2000-087220/07.  
XX PT Novel nucleic acids, used to develop products for the diagnosis and  
PT treatment of disorders involving unwanted cell proliferation,  
PT particularly cancers, especially colon cancer  
XX PS Claim 15; Page 402; 469pp; English.  
XX CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from  
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g.  
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
CC can be used for developing agents for the diagnosis and treatment of  
CC disorders involving unwanted cell proliferation, such as neoplasia,  
CC dysplasia or hyperplasia.



CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

XX Sequence 405 BP; 99 A; 63 C; 76 G; 146 T; 21 other;

XX Query Match 2.5%; Score 131; DB 16; Length 405;

XX Best Local Similarity 100.0%; Pred. No. 5e-31;

XX Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4912 ccccttggtatcttaccatagtttactctggtggacccttaattctcagagtgctaa 4971  
|||||

Db 132 ccccttggtatcttaccatagtttactctggtggacccttaattctcagagtgctaa 191  
|||||

Qy 4972 attctgcccattacaccagagatgctctgatagagacacacatgcaaatgtga 5031  
|||||

Db 192 attctgcccattacaccagagatgctctgatagagacacacatgcaaatgtga 251  
|||||

Qy 5032 aatagtctga 5042  
|||||

Db 252 aatagtctga 262

RESULT 9

AAC06611 ID AAC06611 standard; cDNA; 161 BP.

XX AC AAC06611;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10686.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10686; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX mRNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

XX sequences usually correspond mainly to the 3' untranslated region (UTR)

XX of the mRNA because they are often obtained from oligo-dT primed cDNA

XX libraries. Such ESTs are not well suited for isolating cDNA sequences

XX derived from the 5' ends of mRNAs and even in those cases where longer

XX cDNA sequences have been obtained, the full 5' UTR is rarely included.

XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX They are used to obtain upstream regulatory sequences and to design

XX expression and secretion vectors.

SQ Sequence 161 BP; 54 A; 25 C; 43 G; 39 T; 0 other;

XX Query Match 2.3%; Score 120; DB 21; Length 161;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-27;

XX Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 aggggtagaataatttctgtcattcaagaactagaccatgatggaaaatt 104  
|||||

Db 42 aggggtagaataatttctgtcattcaagaactagaccatgatggaaaatt 101  
|||||

Qy 105 acatattccctgggttcaggaaatcatcagataaaatattcagaagagatggtgaga 164  
|||||

Db 102 acatattccctgggttcaggaaatcatcagataaaatattcagaagagatggtgaga 161  
|||||

RESULT 10

AAX12049 ID AAX12049 standard; DNA; 150 BP.

XX AC AAX12049;

XX 30-MAR-1999 (first entry)

XX Human biallelic polymorphic DNA fragment WI-18387b.

XX Polymorphism; biallelic; human; forensic; paternity testing; disease;

XX detection; phenotypic typing; characteristic; infection; hereditary;

XX autoimmune disease; cancer; inflammation; drug; therapy; medication;

XX treatment; marker; ss.

XX Homo sapiens.

XX WO9820165-A2.

XX 14-MAY-1998.

XX 05-NOV-1997; 97WO-US20313.

XX 06-NOV-1996; 96US-0030455.

XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

XX New isolated nucleic acid segments from the human genome - used for

XX determining polymorphic forms for use in e.g. forensics, paternity

XX testing or phenotypic typing for disease

XX Claim 1; Page 215; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain biallelic

XX polymorphic markers which have been isolated using the primers

XX represented in AAX09121-X10268. The base occupying the polymorphic site

XX is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments

XX can be used in methods for determining polymorphic forms in an individual

XX for use in e.g. forensics, paternity testing or for phenotypic typing for

XX diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan

XX syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,

XX familial hypercholesterolemia, polycystic kidney disease, hereditary

XX spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary

XX haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos

XX syndrome, osteogenesis imperfecta, acute intermittent porphyria,

XX autoimmune diseases, inflammation, cancer, diseases of the nervous

XX system, infection by pathogenic microorganisms, and characteristics such

XX as longevity, appearance (e.g. baldness, obesity), strength, speed,

XX endurance, fertility, and susceptibility or receptivity to particular

XX drugs or therapeutic treatments. The isolated polymorphic nucleic acid

XX segments can also be used to produce medicaments for the treatment or

XX prophylaxis of such diseases.



QY 347 agctccttacacatccct 365  
Db 328 agctccttacacatccct 346

## RESULT 7

AAZ14459  
ID AAZ14459 standard; cDNA; 284 BP.

AC AAZ14459;

XX 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:1928.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX W0938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX Claim 1; Page 1096; 2479pp; English.

XX The present invention describes a library of human polynucleotides  
XX comprising the sequences given in AAZ12532 to AAZ1779. Also described is  
XX a method of detecting differentially expressed genes correlated with the  
XX cancerous state of a mammalian cell, comprising detecting at least one  
XX differentially expressed gene product in a test sample from a cell  
XX suspected of being cancerous, where the gene product is encoded by one  
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The  
XX polynucleotides can be used as a source of primers and probes, which can  
XX be used for a variety of purpose, e.g. detection of expression levels,  
XX mapping, tissue typing or profiling, forensics, genetic analysis and  
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides  
XX can be used for raising antibodies for experimental, diagnostic and  
XX therapeutic purposes. The polynucleotides may also be used to construct  
XX arrays for diagnostics (which may be used to determine function of an  
XX encoded protein); and to detect differences in expression levels between  
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to  
XX identify a genetic predisposition or susceptibility to a disease such as  
XX cancer). The polynucleotides of the invention are especially used in the  
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,  
XX and lung cancer. The polynucleotides can also be used to screen for  
XX peptide analogues and antagonists.

XX Sequence 284 BP; 89 A; 43 C; 46 G; 103 T; 3 other;

Query Match 4.2%; Score 219; DB 20; Length 284;

Best Local Similarity 100.0%; Pred. No. 1.6e-57;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4970 aaatgtctgccattacacacagaagatgctctgtatagagacacacatgcaaatgt 5029  
Db 1 aaatgtctgccattacacacagaagatgctctgtatagagacacacatgcaaatgt 60

QY 5030 gaaatagctcctgaagttcttgattacttacacctcagattgtgcccagaattt 5089  
Db 61 gaaatagctcctgaagttcttgattacttacacctcagattgtgcccagaattt 120

QY 5090 tctggcctttcatggcaatgaaatttttaagaagaagatttaaaagtattttattttaa 5149  
Db 121 tctggcctttcatggcaatgaaatttttaagaagaagatttaaaagtattttattttaa 180

QY 5150 agagtgtgtataaaataatgtactgaattctttatccc 5188

Db 181 agagtgtgtataaaataatgtactgaattctttatccc 219

## RESULT 8

AAT19926

ID AAT19926 standard; cDNA to mRNA; 405 BP.

XX AC AAT19926;

XX 17-JUL-1996 (first entry)

XX Human gene signature HUMGS01054.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

XX OS Homo sapiens.

XX PN W09514772-A1.

XX PD 01-JUN-1995.

XX PF 11-NOV-1994; 94WO-JP01916.

XX PR 12-NOV-1993; 93JP-0355504.

XX PA (MATS/) MATSUBARA K.

XX PA (OKUBO/) OKUBO K.

XX PI Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
XX for diagnosis of abnormal cell function, by preparing cDNA that  
XX reflects relative abundance of corresp. mRNA in specific human  
XX tissues

XX Claim 1; Page 510-511; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX given in AAT19001-T26837 and which is able to hybridise to part of  
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
XX sequences were obtained from 3'-directed cDNA libraries prepared  
XX from various human tissues; synthesis of cDNA was initiated from the  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX untranslated sequence is unique to a particular mRNA species, almost  
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
XX is constructed so as to reflect accurately the relative abundance of  
XX different mRNAs in the particular tissue from which it was derived.  
XX The appearance frequency of a given GS in a cDNA library can be  
XX determined (esp. using primers and probes derived from the GS



PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG02811.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 XX mRNAs encoding secreted proteins. An ORF has been identified within the  
 XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+  
 XX RNAs derived from 30 different tissues. EST sequences usually correspond  
 XX mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 XX well suited for isolating cDNA sequences derived from the 5' ends of  
 XX mRNAs and even in those cases where longer cDNA sequences have been  
 XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 XX mRNAs with intact 5' ends and can therefore be used to obtain full length  
 XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 XX gene therapy and chromosome mapping procedures. They are used to obtain  
 XX upstream regulatory sequences and to design expression and secretion  
 XX vectors.

XX Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

Query Match 5.5%; Score 292; DB 21; Length 295;

Best Local Similarity 100.0%; Pred. No. 1.6e-79;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2562 gttcgtggtcacttggaataaataatcacagtaaatcagaactctaccttaaga 2621

Db 1 gttcgtggtcacttggaataaataatcacagtaaatcagaactctaccttaaga 60

Qy 2622 ttgtcaacaacaatttgcattgtagtgagacttgacagacaggggaaattagtaaa 2681

Db 61 ttgtcaacaacaatttgcattgtagtgagacttgacagacaggggaaattagtaaa 120

Qy 2682 ccagatatgtcacgtctgagacttgctgctgggagtgctattgtgaagctggcacaagaa 2741

Db 121 ccagatatgtcacgtctgagacttgctgctgggagtgctattgtgaagctggcacaagaa 180

Qy 2742 cccgtttaccatgaataatcacattagacaataatcagctatgtgctattgctatcaac 2801

Db 181 cccgtttaccatgaataatcacattagacaataatcagctatgtgctattgctatcaac 240

Qy 2802 gatgaatgctatcaagtaagaacaagtgttgcacagaacttcacaaaggcc 2853

Db 241 gatgaatgctatcaagtaagaacaagtgttgcacagaacttcacaaaggcc 292

RESULT 6

AAZ42861

ID AAZ42861 standard; cDNA; 351 BP.

XX AC AAZ42861;

XX 01-FEB-2000 (first entry)

XX Human 5' EST isolated from a cDNA library SEQ ID NO:620.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;  
 XX gene therapy; chromosome mapping; upstream regulatory sequence;  
 XX forensic; location; development; protein synthesis; stability;

regulation; identification; ss.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-IB00712.

XX 09-APR-1998; 98US-0057719.

XX 28-APR-1998; 98US-0069047.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-038446/03.

XX P-PSDB; AAY65247.

XX Novel secreted protein 5' expressed sequence tag sequences used in  
 XX diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX Claim 1; Page 495; 837pp; English.

XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)  
 XX sequences, corresponding to human secreted proteins. AAY64651 to  
 XX AAY65438 represent the EST-related proteins corresponding to AAZ42265 to  
 XX AAZ43052. The 5' ESTs can be used for producing secreted human gene  
 XX products. They can be used to identify and isolate 5' untranslated  
 XX regions (UTRs) and upstream regulatory regions which control the  
 XX location, development stage, rate, and quantity of protein synthesis, as  
 XX well as stability of mRNA. The ESTs are also useful as probes for  
 XX chromosome mapping, and to obtain full length cDNA clones. The ESTs can  
 XX also be used in forensic procedures to identify individuals, or in  
 XX diagnostic procedures to identify individuals having genetic diseases  
 XX resulting from abnormal gene expression. The products may also be used in  
 XX gene therapy protocols. The nucleic acids encoding signal peptides can be  
 XX used for directing extracellular secretion of a polypeptide or the  
 XX insertion of a polypeptide into a membrane, or importing a polypeptide  
 XX into a cell. The proteins encoded by the EST sequences may be useful in  
 XX treating a variety of human conditions. Secreted proteins have  
 XX therapeutic value, and the identification of new secreted proteins is  
 XX valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent  
 XX sequences used in the exemplification of the present invention.

XX Sequence 351 BP; 107 A; 63 C; 80 G; 99 T; 2 other;

Query Match 5.1%; Score 268; DB 21; Length 351;

Best Local Similarity 99.7%; Pred. No. 2.6e-72;

Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 47 gggtagaataatttctgctcattcattcaagaactaggaccattgatgaaaaattac 106

Db 28 gggtagaataatttctgctcattcattcaagaactaggaccattgatgaaaaattac 87

Qy 107 atatccgctgggtcgaaggaaatcatcagataaaatctctaaaggagagtggtgagacg 166

Db 88 atatccgctgggtcgaaggaaatcatcagataaaatctctaaaggagagtggtgagacg 147

Qy 167 attaaagatggttgaaacttttctgatatgaccagactctgagaagaaagga 226

Db 148 attaaagatggttgaaacttttctgatatgaccagactctgagaagaaagga 207

Qy 227 gctttattaaacctagctttacatcttctgcttcagatttttctcaagcatcctggttaa 286

Db 208 gctttattaaacctagctttacatcttctgcttcagatttttctcaagcatcctggttaa 267

Qy 287 agatgttcgcttactgtagctgctgctgctgctgctgctgctgctgctgctgctgctg 346

Db 268 agatgttcgcttactgtagctgctgctgctgctgctgctgctgctgctgctgctgctg 327



QY 3290 catcatgtcaagagactactacatacagtttggaattctcttaagaccggtactaccagc 3349  
 DB 22 catcatgtcaagagactactacatacagtttggaattctcttaagaccggtactaccagc 81  
 QY 3350 tcgtttcttcactcaactgacaagaatttcagtaacacacaaaattatctgcctctga 3409  
 DB 82 tcgtttcttcactcaactgacaagaatttcagtaacacacaaaattatctgcctctga 141  
 QY 3410 aatgaatacatitttctcactcctggaaacctaaacacaaatgttctagggagctgttaa 3469  
 DB 142 aatgaatacatitttctcactcctggaaacctaaacacaaatgttctagggagctgttaa 201  
 QY 3470 caagccacttctcagcagcgaagcaatctcagaccacaaatcagcaatggaactgt 3529  
 DB 202 caagccacttctcagcagcgaagcaatctcagaccacaaatcagcaatggaactgt 261  
 QY 3530 aagcaatgcaagcagcgtcaaatccaagctctctggaagaataaaaggagagcttga 3589  
 DB 262 aagcaatgcaagcagcgtcaaatccaagctctctggaagaataaaaggagagcttga 321  
 QY 3590 tagttctgaaatggatcacagtgaaatgaagattacacaatgtcttcacctttgcgcgg 3649  
 DB 322 tagttctgaaatggatcacagtgaaatgaagattacacaatgtcttcacctttgcgcgg 381  
 QY 3650 gaaaaaagtgacaagagagagactctgatcttgaaggtctgaat 3696  
 DB 382 gaaaaaagtgacaagagagagactctgatcttgaaggtctgaat 428

## RESULT 4

AAZ15259  
 ID AAZ15259 standard; cDNA; 738 BP.

XX AAZ15259;

DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:2728.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO9338972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX

PS Claim 1; Page 1324; 2479pp; English.

XX

CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.

SQ Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;

Query Match 6.6%; Score 348; DB 20; Length 738;

Best Local Similarity 100.0%; Pred. No. 1.7e-96;

Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3890 agaaaatgaagtgaacagaatagtcgcccaaaaagggtataaagagcgccaccacaaa 3949

DB 51 agaaaatgaagtgaacagaatagtcgcccaaaaagggtataaagagcgccaccacaaa 110

QY 3950 acctcttggtgaggtacacacaaagaagagccaaatgaaactctcaaaaaggagag 4009

DB 111 acctcttggtgaggtacacacaaagaagagccaaatgaaactctcaaaaaggagag 170

QY 4010 caaaaaaatctgacctccagccagagagagagagagagagagagagagagagagag 4069

DB 171 caaaaaaatctgacctccagccagagagagagagagagagagagagagagagagag 230

QY 4070 aaatcggaaacagaagtcacaaagcaaacagacacagagagagagagagagagagag 4129

DB 231 aaatcggaaacagaagtcacaaagcaaacagacacagagagagagagagagagagag 290

QY 4130 agcagaatctcctgaatctagtgcaattgaattccacacagtcacacacagagagag 4189

DB 291 agcagaatctcctgaatctagtgcaattgaattccacacagtcacacacagagagag 350

QY 4190 aggaagaccatcaaaaagccatccatccatccatccatccatccatccatccatccat 4237

DB 351 aggaagaccatcaaaaagccatccatccatccatccatccatccatccatccatccat 398

## RESULT 5

AAZ02817

ID AAC02817 standard; cDNA; 295 BP.

XX AAC02817;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 2815.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

OS

XX EP1033401-A2.

XX 06-SEP-2000.

XX











CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is  
CC located on chromosome 13 at position 13q12-13q. AS3 has a role in  
CC inhibiting cell proliferation and use as a marker for the efficient  
CC diagnosis and treatment of prostate cancer. The invention includes AS3  
CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host  
CC cell transfected with the expression vector, and a method for producing  
CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has  
CC cytosolic activity, and acts to suppress cell proliferation. The AS3  
CC gene is useful as a marker for the efficient diagnosis and treatment of  
CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of  
CC antisense agents for sequence specific modulation of gene expression. The  
CC AS3 protein may be used in the treatment of disorders caused by aberrant  
CC modification or mutation of a gene encoding an AS3 protein, misregulation  
CC of the AS3 gene or aberrant post-translational modification of the AS3  
CC protein. This sequence represents the human AS3 cDNA sequence with an  
CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when  
CC compared with the claimed AS3 cDNA sequence AAA28051.

XX Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 5271; DB 21; Length 5355;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccggagagcccgagtgagcgagtagcgagtcgcaaccgaggggtagaataattt 60  
DB |  
DB 85 ccggagagcccgagtgagcgagtagcgagtcgcaaccgaggggtagaataattt 144  
QY 61 ctgtcatggtcattcaaaactagggaccattgatgaaaaattacatccgctggg 120  
DB |  
DB 145 ctgtcatggtcattcaaaactagggaccattgatgaaaaattacatccgctggg 204  
QY 121 tcaaggaaatcatgataaaatatactaaagagagagatggtgagacgattaaagatggtg 180  
DB |  
DB 205 tcaaggaaatcatgataaaatatactaaagagagagatggtgagacgattaaagatggtg 264  
QY 181 tgaatactttatgatatgaccagagactctgaagagaaagagagctttatttaaac 240  
DB |  
DB 265 tgaatactttatgatatgaccagagactctgaagagaaagagagctttatttaaac 324  
QY 241 tagctttacatcttgccttcagattttttctcaagcatcctgtaagatgttcgttac 300  
DB |  
DB 325 tagctttacatcttgccttcagattttttctcaagcatcctgtaagatgttcgttac 384  
QY 301 tggtagcctgctgcttgctgatatatttcaggatttatgctcctgaagctccttacacat 360  
DB |  
DB 385 tggtagcctgctgcttgctgatatatttcaggatttatgctcctgaagctccttacacat 444  
QY 361 cccctgataaactaaaggatatttatgtttatacaagacagcttgaagggttagagg 420  
DB |  
DB 445 cccctgataaactaaaggatatttatgtttatacaagacagcttgaagggttagagg 504  
QY 421 atacaagagcccacattcaataggtattttatttacttgagaacattgcttggtca 480  
DB |  
DB 505 atacaagagcccacattcaataggtattttatttacttgagaacattgcttggtca 564  
QY 481 agtcaataacatatgtttgagttagaagatgcaatgaaattttccaccagctataca 540  
DB |  
DB 565 agtcaataacatatgtttgagttagaagatgcaatgaaattttccaccagctataca 624  
QY 541 gaacctattttcagttataaaacattggccacacatcagaagtcacatgacatggtag 600  
DB |  
DB 625 gaacctattttcagttataaaacattggccacacatcagaagtcacatgacatggtag 684  
QY 601 acctatgagctctattttgtgaaggtgatacagtgctcaggagcttttggatacgg 660  
DB |  
DB 685 acctatgagctctattttgtgaaggtgatacagtgctcaggagcttttggatacgg 744  
QY 661 ttttagtaaatctggtaccctgctcaataagaatttaacaagcaagcatatgattggcaa 720  
DB |  
DB 745 ttttagtaaatctggtaccctgctcaataagaatttaacaagcaagcatatgattggcaa 804

QY 721 aggtcttactgaagagacagcagctcaagctatttgagccatattaccacttttttaac 780  
DB |  
DB 805 aggtcttactgaagagacagcagctcaagctatttgagccatattaccacttttttaac 864  
QY 781 aggtcttactgaagagacacatctatcagcgatttgtagagcagcatgttttgaactaa 840  
DB |  
DB 865 aggtcttactgaagagacacatctatcagcgatttgtagagcagcatgttttgaactaa 924  
QY 841 ttttgagagctcaacatattgatagtcatttgctgctctctgttttaccagcttgat 900  
DB |  
DB 925 ttttgagagctcaacatattgatagtcatttgctgctctctgttttaccagcttgat 984  
QY 901 ttaaaattaaagagcaatgataatgagcgccctacaaagtgttcaaaactactggcaaaa 960  
DB |  
DB 985 ttaaaattaaagagcaatgataatgagcgccctacaaagtgttcaaaactactggcaaaa 1044  
QY 961 tgttggggcacaaggattcagaatttggctctcctcaaaacagccactttggcagtgctact 1020  
DB |  
DB 1045 tgttggggcacaaggattcagaatttggctctcctcaaaacagccactttggcagtgctact 1104  
QY 1021 tgggcaggttttaatgatccatgtaccatccgctcgtggaatgtgtgaaatttctagcc 1080  
DB |  
DB 1105 tgggcaggttttaatgatccatgtaccatccgctcgtggaatgtgtgaaatttctagcc 1164  
QY 1081 attgtctcatgaaccatcctcgtatttagcaaaagacttaacagagtatctttaaagtgaagt 1140  
DB |  
DB 1165 attgtctcatgaaccatcctcgtatttagcaaaagacttaacagagtatctttaaagtgaagt 1224  
QY 1141 cacatgacctggaggaagctatttagacatgatttattgtgtcatttagttacagctgcta 1200  
DB |  
DB 1225 cacatgacctggaggaagctatttagacatgatttattgtgtcatttagttacagctgcta 1284  
QY 1201 aaaaagatttctcctggtcgaatgacatcacttaacttaattttgtgagagagagacattag 1260  
DB |  
DB 1285 aaaaagatttctcctggtcgaatgacatcacttaacttaattttgtgagagagagacattag 1344  
QY 1261 acaaacgatgagagtagcgcgaagaagcattgaggaactgcccgaatttataagaagt 1320  
DB |  
DB 1345 acaaacgatgagagtagcgcgaagaagcattgaggaactgcccgaatttataagaagt 1404  
QY 1321 atgctttacagtcagcagctggaaaagatgctgcgaacacagatgagctggatcacaagaca 1380  
DB |  
DB 1405 atgctttacagtcagcagctggaaaagatgctgcgaacacagatgagctggatcacaagaca 1464  
QY 1381 aattgctacatatattatcaaaaatagatttagatgctgactactgttgaacggatct 1440  
DB |  
DB 1465 aattgctacatatattatcaaaaatagatttagatgctgactactgttgaacggatct 1524  
QY 1441 ttgctcaatcatggttctccacaaatttagaaactacagacggatgaaatgcttatatt 1500  
DB |  
DB 1525 ttgctcaatcatggttctccacaaatttagaaactacagacggatgaaatgcttatatt 1584  
QY 1501 acttgatgcccactggatttaaatctgctgaaagcattgaaatgctggaatgctc 1560  
DB |  
DB 1585 acttgatgcccactggatttaaatctgctgaaagcattgaaatgctggaatgctc 1644  
QY 1561 aaaaatcgtccgcacatcaagtaaaagatttgcttgactgatttaagcaaccccaaacag 1620  
DB |  
DB 1645 aaaaatcgtccgcacatcaagtaaaagatttgcttgactgatttaagcaaccccaaacag 1704  
QY 1621 atgcccagtgcaaggccatattttcaaaagtgtggtttattacaagaatttacctgac 1680  
DB |  
DB 1705 atgcccagtgcaaggccatattttcaaaagtgtggtttattacaagaatttacctgac 1764  
QY 1681 ctggttaaggctcaggatttcatgaagaataatcacacaggtgtttagaagatgatgagaaaa 1740  
DB |  
DB 1765 ctggttaaggctcaggatttcatgaagaataatcacacaggtgtttagaagatgatgagaaaa 1824  
QY 1741 taagaagaagcaggttagaagtagctttagtccaaactctcctcgaagcagctgaaggtt 1800  
DB |  
DB 1825 taagaagaagcaggttagaagtagctttagtccaaactctcctcgaagcagctgaaggtt 1884  
QY 1801 gtgctgctgaataactaaagaagtgtgggcaaccccaaacagcctacaaatccttccctgg 1860



Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.  
Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;  
chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;  
diagnosis; treatment; cytostatic; human; ss.  
Homo sapiens.

DE	FT	2397..2555	exon	FT
XX	FT	/tag= t		FT
KW	FT	/number= 21		FT
KW	FT	2556..2624	exon	FT
KW	FT	/tag= u		FT
XX	FT	/number= 22		FT
OS	FT	2625..2761	exon	FT
XX	FT	/tag= v		FT
FH	FT	/number= 23		FT
FT	FT	2672..2885	exon	FT
FT	FT	/tag= w		FT
FT	FT	/number= 24		FT
FT	FT	2886..3090	exon	FT
FT	FT	/tag= x		FT
FT	FT	/number= 25		FT
FT	FT	3091..3205	exon	FT
FT	FT	/tag= y		FT
FT	FT	/number= 26		FT
FT	FT	3206..3338	exon	FT
FT	FT	/tag= z		FT
FT	FT	/number= 27		FT
FT	FT	3339..3458	exon	FT
FT	FT	/tag= aa		FT
FT	FT	/number= 28		FT
FT	FT	3459..3521	exon	FT
FT	FT	/tag= ab		FT
FT	FT	/number= 29		FT
FT	FT	3522..3667	exon	FT
FT	FT	/tag= ac		FT
FT	FT	/number= 30		FT
FT	FT	3668..3773	exon	FT
FT	FT	/tag= ad		FT
FT	FT	/number= 31		FT
FT	FT	3774..4213	exon	FT
FT	FT	/tag= ae		FT
FT	FT	/number= 32		FT
FT	FT	4214..4438	exon	FT
FT	FT	/tag= af		FT
FT	FT	/number= 33		FT
FT	FT	4314..4408	misc_signal	FT
FT	FT	/tag= ag		FT
FT	FT	/note= "Destabilizing signal"		FT
FT	FT	4439..5337	exon	FT
FT	FT	/tag= ah		FT
FT	FT	/number= 34		FT
FT	FT	5312..5317	polyA_signal	FT
FT	FT	/tag= ai		FT
FT	FT	5333..5337	misc_feature	FT
FT	FT	/tag= aj		FT
FT	FT	/note= "Poly adenylation cleavage site"		FT
XX	XX			XX
PN	PN	WO200050454-A1.		PN
XX	XX			XX
PD	PD	31-AUG-2000.		PD
XX	XX			XX
PF	PF	24-FEB-2000; 2000WO-US04732.		PF
XX	XX			XX
PR	PR	24-FEB-1999; 99US-0121461.		PR
XX	XX			XX
PA	PA	(TUFT ) TUFTS COLLEGE.		PA
XX	XX			XX
PI	PI	Soto AM, Sonnenschein C, Geck P, Szelei J;		PI
XX	XX			XX
DR	DR	WPI; 2000-565451/52.		DR
XX	XX	P-PSDB; AAY94702.		XX
PT	PT	New human androgen-induced tumor suppressor cDNA sequence termed		PT
PT	PT	'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient		PT
XX	XX	diagnosis and treatment of prostate cancer -		XX
PS	PS	Example 4; Fig 6; 152pp; English.		PS
XX	XX	This invention relates to a human androgen-induced tumour suppressor		XX







1021 tgggagggttaagtatccatgtaccatccgctggaattgtaaatgtctagcc 1080  
1081 attgctctcaagaaccatccgtattagcaaaagacttaacagagtagtcttaaaagttaggt 1140  
1081 attgctctcaagaaccatccgtattagcaaaagacttaacagagtagtcttaaaagttaggt 1140  
1141 cacatgacctgaggagctatttagacatgagtattgtgtcaatagttcacagctgcta 1200  
1141 cacatgacctgaggagctatttagacatgagtattgtgtcaatagttcacagctgcta 1200  
1201 aaagagatattcttggccaatgagcacttacttacttaattttgagagagaaacattag 1260  
1201 aaagagatattcttggccaatgagcacttacttacttaattttgagagagaaacattag 1260  
1261 acaaacgagtgaggagtagcgcgaagaagccatgagtgaggcttgcccaaatatttaagaat 1320  
1261 acaaacgagtgaggagtagcgcgaagaagccatgagtgaggcttgcccaaatatttaagaat 1320  
1321 atgctttacagtcagcagctggaagagtagtgcgaagaagtagcatggatgcgaagaca 1380  
1321 atgctttacagtcagcagctggaagagtagtgcgaagaagtagcatggatgcgaagaca 1380  
1381 aattgctacatatattatcaaaatagtagtattgtagcacttacttggtaacgagatct 1440  
1381 aattgctacatatattatcaaaatagtagtattgtagcacttacttggtaacgagatct 1440  
1441 ttgctcaatcacatggttctccacaatttagaaactacagacgagtagaattgcttatt 1500  
1441 ttgctcaatcacatggttctccacaatttagaaactacagacgagtagaattgcttatt 1500  
1501 acttgatgccacacagtggaatttaaatgctgtgaaagcattggaatgtggaatgtc 1560  
1501 acttgatgccacacagtggaatttaaatgctgtgaaagcattggaatgtggaatgtc 1560  
1561 aaaaatgctccgacatcaagtaaaagatttgcctgactgtagtaagcaacccaacag 1620  
1561 aaaaatgctccgacatcaagtaaaagatttgcctgactgtagtaagcaacccaacag 1620  
1621 atgcccagtgccaagggcattttccaaaagtgatggttattacaagaattttacctgac 1680  
1621 atgcccagtgccaagggcattttccaaaagtgatggttattacaagaattttacctgac 1680  
1681 ctggttaagctcaggtattcatgaagaatttcaacacaggtgttagaagatgagagaaa 1740  
1681 ctggttaagctcaggtattcatgaagaatttcaacacaggtgttagaagatgagagaaa 1740  
1741 taagaagcaggttagaagtagtctttagtccaaactgctcctgcaagcaggtggaaggtt 1800  
1741 taagaagcaggttagaagtagtctttagtccaaactgctcctgcaagcaggtggaaggtt 1800  
1801 gtgtgctgtaataactaagaaggttggaaccccccaacagcctacaaatccttctctgg 1860  
1801 gtgtgctgtaataactaagaaggttggaaccccccaacagcctacaaatccttctctgg 1860  
1861 aaatgatacagtttctctggagagtagcagcctgtgacatagataccgaatctatca 1920  
1861 aaatgatacagtttctctggagagtagcagcctgtgacatagataccgaatctatca 1920  
1921 gtgctcttataaacaagtgaacaaatcaatagatggaacagcagatgatgaagtgagg 1980  
1921 gtgctcttataaacaagtgaacaaatcaatagatggaacagcagatgatgaagtgagg 1980  
1981 gtgttccaactgatcaagccatcaagcgggttcttgaatgctttaagggtactctcttta 2040  
1981 gtgttccaactgatcaagccatcaagcgggttcttgaatgctttaagggtactctcttta 2040  
2041 cacatcccatctcattctgctggaacatttgaatcattactgctgtgtctgaaaaa 2100  
2041 cacatcccatctcattctgctggaacatttgaatcattactgctgtgtctgaaaaa 2100  
2101 tggatgatgaaaaagtagcagaagctgcactacaaatttccaaaacacaggaagcaaaa 2160  
2101 tggatgatgaaaaagtagcagaagctgcactacaaatttccaaaacacaggaagcaaaa 2160

2101 tggatgatgaaaaagtagcagaagctgcactacaaatttccaaaacacaggaagcaaaa 2160  
2161 ttgaagagattttccacacatcagatcagcctgtgcttctgtttttacatcacaaatctta 2220  
2161 ttgaagagattttccacacatcagatcagcctgtgcttctgtttttacatcacaaatctta 2220  
2221 aaaaaggacccccctgcagccaaatatagcatttcatgtatccatgcgatatcttctta 2280  
2221 aaaaaggacccccctgcagccaaatatagcatttcatgtatccatgcgatatcttctta 2280  
2281 gtaaaagagaccaggtttgcacagatatttgagcctctgcataagagccttagatccaaaga 2340  
2281 gtaaaagagaccaggtttgcacagatatttgagcctctgcataagagccttagatccaaaga 2340  
2341 acctggaacatctcataaacaccattggttactatttggtcatattgcttctcttgcacctg 2400  
2341 acctggaacatctcataaacaccattggttactatttggtcatattgcttctcttgcacctg 2400  
2401 atcaattgctgctctcttggaatcttgggttagctacttcttcatgtgaaagatcttctca 2460  
2401 atcaattgctgctctcttggaatcttgggttagctacttcttcatgtgaaagatcttctca 2460  
2461 tgaatgatcggtctccagggaaaaaacacactaaactttgggttccagatgaaagaagtat 2520  
2461 tgaatgatcggtctccagggaaaaaacacactaaactttgggttccagatgaaagaagtat 2520  
2521 ctctgagacaattggtcaaaattcaggctattaaaaatgatggttctgactcttgga 2580  
2521 ctctgagacaattggtcaaaattcaggctattaaaaatgatggttctgactcttgga 2580  
2581 tgaataataatcacagataaatacaggaaacttctaccttaagattgcttaacaaatattgc 2640  
2581 tgaataataatcacagataaatacaggaaacttctaccttaagattgcttaacaaatattgc 2640  
2641 atagtgatggaacttgacagacagggaaaaatagtaaacacagatatgtcacgctga 2700  
2641 atagtgatggaacttgacagacagggaaaaatagtaaacacagatatgtcacgctga 2700  
2701 gacttgctgctgggagtgctattgtgaagctggcacaagaacctgttaccatgaaatca 2760  
2701 gacttgctgctgggagtgctattgtgaagctggcacaagaacctgttaccatgaaatca 2760  
2761 tcaatagaacaatatcagctatgctatgcttagctatcaacgagatgctatcaagtaa 2820  
2761 tcaatagaacaatatcagctatgctatgcttagctatcaacgagatgctatcaagtaa 2820  
2821 gacaaggtttgcccagaacttcaaaagcctttcccggtttacggttccacttgagt 2880  
2821 gacaaggtttgcccagaacttcaaaagcctttcccggtttacggttccacttgagt 2880  
2881 atagtgcaatctgtgcccctttgtgcaaaagatcctgtaaaggagagagctcatgcta 2940  
2881 atagtgcaatctgtgcccctttgtgcaaaagatcctgtaaaggagagagctcatgcta 2940  
2941 ggcaatggttggtaaaaataaaatgtaagcggagtagtctgaaagcagcatgcagctg 3000  
2941 ggcaatggttggtaaaaataaaatgtaagcggagtagtctgaaagcagcatgcagctg 3000  
3001 ttagtgaataattattgtcttcttaccagagtagtgtgttccatatacaattcaccttt 3060  
3001 ttagtgaataattattgtcttcttaccagagtagtgtgttccatatacaattcaccttt 3060  
3061 tggcacatgaccagatattgtcaaaagtaggtagtattgaaacacttaagaagttaag 3120  
3061 tggcacatgaccagatattgtcaaaagtaggtagtattgaaacacttaagaagttaag 3120  
3121 aatgcttcttgggttcttggaataatttaattggttaaaatgaaataacagctcagctt 3180  
3121 aatgcttcttgggttcttggaataatttaattggttaaaatgaaataacagctcagctt 3180  
3181 ttatcagaagaagtggtagaataatttaaaacaaacaaagaatgcccgaagcagcatgag 3240  
3181 ttatcagaagaagtggtagaataatttaaaacaaacaaagaatgcccgaagcagcatgag 3240







C 961	24	0.5	1671	22	AAF33038	Human secreted pro	FT	exon	174...377
C 962	24	0.5	1673	21	AAZ65064	Membrane-bound pro	FT		/tag= d
C 963	24	0.5	1674	21	AAZ58633	Human PRO836 prote	FT		/number= 3
C 964	24	0.5	1674	22	AAF44210	Human PRO836 (UNQ5	FT	exon	378...464
C 965	24	0.5	1678	21	AAF16140	Human prostate can	FT		/tag= e
C 966	24	0.5	1681	22	AAF33135	Human secreted pro	FT		/number= 4
C 967	24	0.5	1688	15	AAQ64613	Human mdr-1 promot	FT	exon	465...562
C 968	24	0.5	1693	20	AAV69206	Canine ICAM-R poly	FT		/tag= f
C 969	24	0.5	1693	21	AAZ24362	Canine ICAM-R DNA	FT		/number= 5
C 970	24	0.5	1700	21	AAZ52505	Human secreted pro	FT	exon	563...689
C 971	24	0.5	1704	22	AAC91180	ADAM gene #4. Hom	FT		/tag= e
C 972	24	0.5	1705	16	AAQ79621	Schizosaccharomyce	FT		/number= 6
C 973	24	0.5	1705	22	AAF44264	Human PRO3301 nucl	FT	exon	690...854
C 974	24	0.5	1708	22	AAF26570	DNA encoding human	FT		/tag= f
C 975	24	0.5	1720	22	AAV84533	Human secreted pro	FT		/number= 7
C 976	24	0.5	1724	21	AAA16694	Human secreted pro	FT	exon	855...911
C 977	24	0.5	1733	19	AAV20607	Rat kynurenine-3-h	FT		/tag= g
C 978	24	0.5	1733	20	AAV81394	Human tumour antig	FT		/number= 8
C 979	24	0.5	1738	21	AAZ35178	Corn delta-12 deha	FT	exon	912...1027
C 980	24	0.5	1738	21	AAC99089	Human pancreatic c	FT		/tag= h
C 981	24	0.5	1751	21	AAZ68122	Human secreted pro	FT		/number= 9
C 982	24	0.5	1752	21	AAZ33348	Human secreted pro	FT	exon	1028...1122
C 983	24	0.5	1756	21	AAZ18203	Lung cancer associ	FT		/tag= i
C 984	24	0.5	1758	22	AAF32703	Human secreted pro	FT		/number= 10
C 985	24	0.5	1762	20	AAZ58601	Human leukaemia-as	FT	exon	1123...1268
C 986	24	0.5	1767	20	AAZ51749	DNA encoding a hum	FT		/tag= j
C 987	24	0.5	1770	14	AAQ37684	P.falciparum antig	FT		/number= 11
C 988	24	0.5	1770	14	AAQ37124	P.falciparum antig	FT	exon	1269...1420
C 989	24	0.5	1772	21	AAF21809	Human breast and o	FT		/tag= k
C 990	24	0.5	1776	21	AAA9177	CDNA encoding huma	FT		/number= 12
C 991	24	0.5	1781	14	AAQ46991	CDNA clone encodin	FT	exon	1421...1534
C 992	24	0.5	1781	19	AAV56348	Human ICAM-R CDNA.	FT		/tag= l
C 993	24	0.5	1781	19	AAV54826	CDNA encoding huma	FT		/number= 13
C 994	24	0.5	1781	19	AAV11657	Human ICAM-R CDNA.	FT	exon	1535...1616
C 995	24	0.5	1781	19	AAV19328	Human ICAM-4 CDNA.	FT		/tag= m
C 996	24	0.5	1781	20	AAZ21855	Human ICAM-R codin	FT		/number= 14
C 997	24	0.5	1781	20	AAV69125	Intercellular adhe	FT	exon	1617...1665
C 998	24	0.5	1781	21	AAA97090	Human CDNA encodin	FT		/tag= n
C 999	24	0.5	1781	21	AAA08236	Human ICAM-R prote	FT	exon	1666...1805
C1000	24	0.5	1781	21	AAZ24262	Human ICAM-R CDNA.	FT		/tag= o
							FT	exon	/number= 16
							FT		1806...1921
							FT		/tag= p
							FT	exon	/number= 17
							FT		1922...2027
							FT		/tag= q
							FT		/number= 18
							FT	exon	2028...2188
							FT		/tag= r
							FT	exon	/number= 19
							FT		2189...2312
							FT		/tag= s
							FT		/number= 20
							FT	exon	2313...2471
							FT		/tag= t
							FT	exon	/number= 21
							FT		2472...2540
							FT		/tag= u
							FT		/number= 22
							FT	exon	2541...2677
							FT		/tag= v
							FT	exon	/number= 23
							FT		2678...2801
							FT		/tag= w
							FT	exon	/number= 24
							FT		2802...3006
							FT		/tag= x
							FT	exon	/number= 25
							FT		3007...3121
							FT		/tag= y
							FT		/number= 26
							FT	exon	3122...3254

## ALIGNMENTS

RESULT 1  
 AAA28051  
 ID AAA28051 standard; cDNA; 5271 BP.  
 XX  
 AC AAA28051;  
 XX  
 DT 01-DRC-2000 (first entry)  
 XX Human androgen shutoff gene 3 (AS3) cDNA sequence.  
 DE  
 XX Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;  
 KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;  
 KW diagnosis; treatment; cytostatic; human; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT exon 1..46  
 FT /tag= a  
 FT /number= 1  
 FT 47..173  
 FT /tag= b  
 FT /number= 2  
 FT 66..4241  
 FT /tag= c  
 FT /product= "AS3"  
 FT /note= "Androgen shutoff gene 3 protein, the CDS is  
 specifically claimed as SEQ ID #3".



c 815	24	0.5	1257	13	AAQ24292	(1-3,1-4)-beta-glu	c 888	24	0.5	1462	20	AA241241	Human normal ovari
c 816	24	0.5	1260	21	AA250084	Impatiens balsamia	c 889	24	0.5	1469	14	AAQ40323	Sequence of the CD
c 817	24	0.5	1270	16	AAQ87101	The Tpa gene Syn	c 890	24	0.5	1475	21	AA259122	Human secreted pro
c 818	24	0.5	1270	18	AA263685	Full-length Tpa-1	c 891	24	0.5	1475	21	AA259122	Human normal pancr
c 819	24	0.5	1270	20	AA241310	Human normal ovari	c 892	24	0.5	1481	21	AA269406	Human secreted pro
c 820	24	0.5	1270	20	AA241310	Human normal anioni	c 893	24	0.5	1481	21	AA297081	Human secreted pro
c 821	24	0.5	1276	17	AA242071	Human nuclear prol	c 894	24	0.5	1493	13	AAQ28411	Glia activating fa
c 822	24	0.5	1276	21	AA279777	Human secreted pro	c 895	24	0.5	1493	13	AAQ27793	Human TANGO 295 co
c 823	24	0.5	1278	21	AA21698	Human breast and o	c 896	24	0.5	1500	21	AA253986	Arabidopsis thalia
c 824	24	0.5	1278	22	AA27543	Caenorhabditis ele	c 897	24	0.5	1508	22	AA27025	Human PRO1773 (UNQ
c 825	24	0.5	1282	21	AA293379	Human secreted pro	c 898	24	0.5	1508	22	AA292067	Human PRO1773 CDNA
c 826	24	0.5	1284	20	AA252938	Human prostate tum	c 899	24	0.5	1508	22	AA254207	DNA encoding prote
c 827	24	0.5	1284	20	AA221094	Human secreted pro	c 900	24	0.5	1509	19	AA259679	Human secreted pro
c 828	24	0.5	1293	21	AA293129	Human secreted pro	c 901	24	0.5	1516	21	AA269497	Human secreted pro
c 829	24	0.5	1295	21	AA261171	Regulatory sequence	c 902	24	0.5	1516	21	AA259440	Nucleotide sequenc
c 830	24	0.5	1297	13	AAQ21645	3' coding sequence	c 903	24	0.5	1520	8	AA270223	Plasmid pBRV2 inse
c 831	24	0.5	1299	21	AA216666	Human secreted pro	c 904	24	0.5	1520	8	AA271320	Sequence encoding
c 832	24	0.5	1300	18	AA264848	Meripilus giganteu	c 905	24	0.5	1521	7	AA260937	Plasmid pBRV2 inse
c 833	24	0.5	1300	21	AA281109	Human secreted pro	c 906	24	0.5	1522	19	AA264309	Human secreted pro
c 834	24	0.5	1300	21	AA281109	Human secreted pro	c 907	24	0.5	1522	19	AA284835	Nucleotide sequenc
c 835	24	0.5	1302	20	AA290446	Human secreted pro	c 908	24	0.5	1524	21	AA237027	Human PRO1478 (UNQ
c 836	24	0.5	1303	20	AA242221	Human normal bladd	c 909	24	0.5	1524	22	AA254212	DNA encoding prote
c 837	24	0.5	1305	19	AA262770	Trichomonas vagina	c 910	24	0.5	1525	13	AAQ31200	Myo-inositol O-met
c 838	24	0.5	1305	19	AA262770	Human normal ovari	c 911	24	0.5	1525	18	AA295213	cDNA encoding the
c 839	24	0.5	1315	20	AA241312	Human normal ovari	c 912	24	0.5	1525	19	AA222704	Mouse recombinase
c 840	24	0.5	1320	16	AA229190	Glycoprotein 39 3'	c 913	24	0.5	1525	21	AA227911	cDNA encoding IL-1
c 841	24	0.5	1320	19	AA220948	Coffee-fruit speci	c 914	24	0.5	1525	21	AA208554	Murine REC2 serine
c 842	24	0.5	1326	20	AA234867	cDNA encoding Brus	c 915	24	0.5	1525	22	AA202334	Murine interleukin
c 843	24	0.5	1328	19	AA258754	Human secreted pro	c 916	24	0.5	1525	22	AA281415	Mouse IL-13 recept
c 844	24	0.5	1334	21	AA252490	Human MN gene intr	c 917	24	0.5	1529	21	AA259838	Human secreted pro
c 845	24	0.5	1334	21	AA216571	Human MN intron 6	c 918	24	0.5	1530	8	AA270222	Plasmid pBRG4 inse
c 846	24	0.5	1338	20	AA297959	Human secreted pro	c 919	24	0.5	1531	7	AA260936	Plasmid pBRG4 inse
c 847	24	0.5	1338	21	AA276985	Human ORFX ORF2540	c 920	24	0.5	1531	21	AA297369	Human colorectal c
c 848	24	0.5	1341	20	AA257178	Human KCNQ3 intron	c 921	24	0.5	1533	20	AA284490	Human secreted pro
c 849	24	0.5	1342	21	AA226424	Human secreted pro	c 922	24	0.5	1533	22	AA232702	Human secreted pro
c 850	24	0.5	1348	15	AAQ78278	T. niveum GAPDH	c 923	24	0.5	1534	8	AA271319	Sequence encoding
c 851	24	0.5	1348	15	AAQ50401	Arabidopsis thalia	c 924	24	0.5	1535	21	AA216550	Human breast and o
c 852	24	0.5	1353	22	AA291863	Human secreted pro	c 925	24	0.5	1537	21	AA279967	Human secreted pro
c 853	24	0.5	1353	9	AA280457	Sequence of cDNA e	c 926	24	0.5	1538	20	AA284593	Human secreted pro
c 854	24	0.5	1355	20	AA207164	Human lung tumour	c 927	24	0.5	1538	21	AA208416	Human tumour suppr
c 855	24	0.5	1355	21	AA279076	Human lung tumour-	c 928	24	0.5	1539	20	AA223317	Human PDSP-1 DNA
c 856	24	0.5	1356	20	AA297957	Human secreted pro	c 929	24	0.5	1547	20	AA233448	Oryza sativa L. pi
c 857	24	0.5	1358	22	AA285033	Soybean HES1 homol	c 930	24	0.5	1547	21	AA259424	Human secreted pro
c 858	24	0.5	1359	21	AA266430	Human secreted pro	c 931	24	0.5	1549	21	AA277188	Human ORFX ORF2743
c 859	24	0.5	1360	21	AA299029	Human TGC839 nucle	c 932	24	0.5	1551	19	AA203028	Globodera rostochl
c 860	24	0.5	1371	19	AA233755	Mouse neutrophil e	c 933	24	0.5	1554	20	AA216675	Xenopus WA345 prot
c 861	24	0.5	1374	21	AA239407	Corn SR2 homology	c 934	24	0.5	1558	18	AA222942	Human cathepsin y
c 862	24	0.5	1376	15	AAQ56747	Ryegrass Lol pv al	c 935	24	0.5	1559	21	AA247168	Mouse sphingosine
c 863	24	0.5	1377	21	AAQ55190	Human secreted pro	c 936	24	0.5	1571	21	AA236366	Arabidopsis thalia
c 864	24	0.5	1378	19	AA259706	Human secreted pro	c 937	24	0.5	1573	19	AA243607	Human secreted pro
c 865	24	0.5	1381	21	AA248572	Soybean protein ph	c 938	24	0.5	1573	21	AA254286	Human mesenchymal
c 866	24	0.5	1386	19	AA211855	Homo sapiens Tub I	c 939	24	0.5	1574	21	AA265325	Human secreted pro
c 867	24	0.5	1386	21	AA277185	Human ORFX ORF2740	c 940	24	0.5	1580	21	AA261288	Human secreted pro
c 868	24	0.5	1387	22	AA202201	Chondrus crispus s	c 941	24	0.5	1589	21	AA295198	Human UGT2B4 exon
c 869	24	0.5	1391	17	AA207329	Human calcium sign	c 942	24	0.5	1592	21	AA277050	Human ORFX ORF2605
c 870	24	0.5	1396	21	AA298027	Human secreted pro	c 943	24	0.5	1607	18	AA278587	Malassezia fungus
c 871	24	0.5	1404	20	AA269207	DNA sequence of ca	c 944	24	0.5	1610	20	AA252902	Human prostate tum
c 872	24	0.5	1404	21	AA224363	Canine ICAM-R DNA	c 945	24	0.5	1614	21	AA221791	Human breast and o
c 873	24	0.5	1405	20	AA280908	Human cDNA clone H	c 946	24	0.5	1616	21	AA252506	Human secreted pro
c 874	24	0.5	1407	7	AA260374	Sequence encoding	c 947	24	0.5	1623	19	AA207059	Mouse mitogen acti
c 875	24	0.5	1407	14	AAQ52515	Sequence encoding	c 948	24	0.5	1624	19	AA209669	Human cathepsin K
c 876	24	0.5	1407	15	AAQ71647	Human phospholipas	c 949	24	0.5	1624	20	AA284570	Human secreted pro
c 877	24	0.5	1407	18	AA289675	Human phospholipas	c 950	24	0.5	1631	18	AA272795	Beta-1 integrin mo
c 878	24	0.5	1408	21	AA221863	Human breast and o	c 951	24	0.5	1637	20	AA284578	Human secreted pro
c 879	24	0.5	1410	21	AA298057	Human colon cancer	c 952	24	0.5	1637	21	AA277959	Human cancer associ
c 880	24	0.5	1427	20	AA277561	Human ovarian tumo	c 953	24	0.5	1637	22	AA275093	Human colon associ
c 881	24	0.5	1428	20	AA224403	Maize myo-inositol	c 954	24	0.5	1642	19	AA240507	Human secreted acti
c 882	24	0.5	1438	21	AA293098	Thiosulphate sulph	c 955	24	0.5	1643	20	AA207062	Mouse mitogen acti
c 883	24	0.5	1446	21	AA277994	Human cancer assoc	c 956	24	0.5	1647	21	AA213841	Aspergillus fumiga
c 884	24	0.5	1452	21	AA270123	Plasmodium falci	c 957	24	0.5	1651	21	AA215694	Cancer suppressor
c 885	24	0.5	1457	21	AA276037	Human ORFX ORF1592	c 958	24	0.5	1651	22	AA297914	Human secreted pro
c 886	24	0.5	1457	21	AA234553	cDNA encoding huma	c 959	24	0.5	1670	19	AA246310	Human secreted pro
c 887	24	0.5	1459	21	AA277423	Human ORFX ORF2978	c 960	24	0.5	1670	20	AA284836	cDNA molecule prod



c 669	0.5	867	18	AA66791	Partial seq. obtai	742	24	0.5	1020	21	AA49558	Partial DNA sequen
c 670	0.5	870	17	AA32400	Flea allergen cDNA	743	24	0.5	1021	21	AA297100	Human secreted pro
c 671	0.5	874	20	AA22219	Human secreted pro	c 744	24	0.5	1024	21	AA25777	Human ORFX ORF1332
c 672	0.5	877	19	AA04635	Human cytoplasmic	745	24	0.5	1024	21	AA297206	Human prostate can
c 673	0.5	878	21	AA294097	Haematopoietic ste	746	24	0.5	1026	21	AA277906	Human cancer assoc
c 674	0.5	884	20	AA241356	Human normal uterus	c 747	24	0.5	1027	20	AA284690	CBCAL005 gene. Ho
c 675	0.5	885	21	AA278135	Human cancer assoc	c 748	24	0.5	1027	21	AA298781	Human prostate can
c 676	0.5	888	21	AA252552	Human secreted pro	c 749	24	0.5	1029	21	AA215949	Human prostate can
c 677	0.5	891	20	AA206233	Human secreted pro	c 750	24	0.5	1031	20	AA224419	Human bladder tumo
c 678	0.5	893	20	AA208308	Human lung tumour-	c 751	24	0.5	1033	17	AA236363	High mobility grou
c 679	0.5	893	21	AA279160	Human lung tumour-	c 752	24	0.5	1033	21	AA289462	Corn leud subunit
c 680	0.5	893	22	AA26575	DNA encoding human	c 753	24	0.5	1038	20	AA227247	Human secreted pro
c 681	0.5	894	21	AA264640	Partial sequence M	c 754	24	0.5	1040	16	AA292530	P. communis (pear)
c 682	0.5	896	20	AA241976	Human myometrium t	c 755	24	0.5	1040	20	AA258275	Zea mays SCLBr pro
c 683	0.5	899	20	AA215217	Human gene express	c 756	24	0.5	1040	21	AA296932	Human secreted pro
c 684	0.5	900	21	AA279020	Human secreted pro	c 757	24	0.5	1050	22	AA286483	Maize Mac2 coding
c 685	0.5	905	19	AA269627	Human secreted pro	c 758	24	0.5	1054	21	AA295529	Human secreted pro
c 686	0.5	907	19	AA259695	Human secreted pro	c 759	24	0.5	1060	20	AA284447	Mouse brain CNG-4
c 687	0.5	907	20	AA235739	DNA encoding a pr	c 760	24	0.5	1064	16	AA287093	Tpa-1 phage insert
c 688	0.5	911	8	AA271002	Sequence encoding	c 761	24	0.5	1064	18	AA263677	Peroxidase gene, T
c 689	0.5	919	20	AA207555	Maize glutathione-	c 762	24	0.5	1064	20	AA203962	Anionic tobacco pe
c 690	0.5	919	21	AA276260	Maize glutathione-	c 763	24	0.5	1069	20	AA284474	Human secreted pro
c 691	0.5	919	21	AA294795	Maize class II glu	c 764	24	0.5	1071	20	AA220414	Human secreted pro
c 692	0.5	920	21	AA259357	Human secreted pro	c 765	24	0.5	1074	22	AA223934	Human secreted pro
c 693	0.5	925	21	AA252534	Human secreted pro	c 766	24	0.5	1082	20	AA241967	Human myometrium t
c 694	0.5	926	21	AA259967	Human secreted pro	c 767	24	0.5	1083	20	AA210659	CDNA encoding a hu
c 695	0.5	926	22	AA230548	Soybean FLF MADS-b	c 768	24	0.5	1097	22	AA202653	Tomato macrocalyx
c 696	0.5	928	20	AA220207	Soybean raffinose	c 769	24	0.5	1106	21	AA229167	Rice 4-hydroxyphen
c 697	0.5	928	20	AA261395	DNA encoding a hum	c 770	24	0.5	1109	22	AA200288	M. cinnamomea xylo
c 698	0.5	928	21	AA274443	Human secreted pro	c 771	24	0.5	1111	20	AA252949	Human prostate tum
c 699	0.5	928	21	AA269576	Human secreted pro	c 772	24	0.5	1116	21	AA246246	Amaranthus viridis
c 700	0.5	943	21	AA279888	Human secreted pro	c 773	24	0.5	1117	17	AA273708	AR192 fragment enc
c 701	0.5	943	21	AA298022	Human secreted pro	c 774	24	0.5	1118	21	AA252547	Human secreted pro
c 702	0.5	944	21	AA297113	Human secreted pro	c 775	24	0.5	1122	21	AA259020	Human secreted pro
c 703	0.5	946	21	AA262694	Human secreted pro	c 776	24	0.5	1126	21	AA259409	Human secreted pro
c 704	0.5	947	22	AA259084	Mouse MLP nucleoti	c 777	24	0.5	1128	18	AA286027	Meripilus giganteu
c 705	0.5	952	21	AA278431	Human secreted pro	c 778	24	0.5	1134	21	AA276265	Maize glutathione-
c 706	0.5	954	20	AA237477	Human secreted pro	c 779	24	0.5	1138	22	AA289725	Maize ZmGnsN1-3 gl
c 707	0.5	959	21	AA262630	Human secreted pro	c 780	24	0.5	1140	22	AA265758	DNA encoding human
c 708	0.5	960	21	AA261884	cDNA encoding a hu	c 781	24	0.5	1141	15	AA259437	Human secreted pro
c 709	0.5	963	21	AA222332	Human secreted pro	c 782	24	0.5	1142	15	AA267320	Fisher rat platele
c 710	0.5	963	22	AA244895	Human breast cance	c 783	24	0.5	1148	22	AA274213	DNA encoding eviro
c 711	0.5	964	20	AA208773	Marmoset sperm-spe	c 784	24	0.5	1164	21	AA265061	Membrane-bound pro
c 712	0.5	968	21	AA298075	Human colon cancer	c 785	24	0.5	1164	22	AA244207	Human PRO511 (UNQ5
c 713	0.5	970	20	AA252871	Human prostate tum	c 786	24	0.5	1167	22	AA264199	Human secreted pro
c 714	0.5	970	21	AA298172	Human colon cancer	c 787	24	0.5	1174	20	AA221516	Progression suppre
c 715	0.5	972	19	AA211784	H. contortus pHC-2	c 788	24	0.5	1174	20	AA261432	DNA encoding a hum
c 716	0.5	973	20	AA221192	zebra mays pathogene	c 789	24	0.5	1175	21	AA273905	Human secreted pro
c 717	0.5	975	19	AA261499	Bak binding protei	c 790	24	0.5	1178	21	AA290469	Human uncoupling p
c 718	0.5	976	18	AA263288	Partial human tryp	c 791	24	0.5	1179	19	AA259761	Human secreted pro
c 719	0.5	976	18	AA263289	Partial human tryp	c 792	24	0.5	1180	20	AA284347	Human Apo-2bcr cDN
c 720	0.5	977	21	AA252526	Human secreted pro	c 793	24	0.5	1187	21	AA275882	Human ORFX ORF1437
c 721	0.5	982	21	AA259197	Human secreted pro	c 794	24	0.5	1190	21	AA278384	Human secreted pro
c 722	0.5	983	19	AA217089	Homo sapiens tubul	c 795	24	0.5	1195	21	AA294943	Human carbohydrate
c 723	0.5	984	21	AA215737	Human prostate can	c 796	24	0.5	1195	21	AA294943	Human secreted pro
c 724	0.5	985	20	AA227250	Human secreted pro	c 797	24	0.5	1200	11	AA205728	Gene encoding glut
c 725	0.5	988	20	AA261428	DNA encoding a hum	c 798	24	0.5	1200	21	AA215723	Human prostate can
c 726	0.5	989	20	AA224820	Human secreted pro	c 799	24	0.5	1200	21	AA254293	Human mesenchymal
c 727	0.5	995	20	AA259354	Human secreted pro	c 800	24	0.5	1212	20	AA276499	Human Wisp-3 prote
c 728	0.5	995	21	AA269107	Human secreted pro	c 801	24	0.5	1212	20	AA276500	Human Wisp-3 prote
c 729	0.5	996	16	AA289606	Human membrane pre	c 802	24	0.5	1213	21	AA277817	Human cancer assoc
c 730	0.5	997	20	AA241269	Human normal ovar	c 803	24	0.5	1215	20	AA200620	Human secreted pro
c 731	0.5	1000	21	AA277786	Human cancer assoc	c 804	24	0.5	1216	21	AA298087	Human colon cancer
c 732	0.5	1001	21	AA257481	Arachidonic acid m	c 805	24	0.5	1219	21	AA223434	cDNA encoding huma
c 733	0.5	1001	21	AA257995	Arachidonic acid m	c 806	24	0.5	1229	22	AA264202	Human secreted pro
c 734	0.5	1001	21	AA257996	Arachidonic acid m	c 807	24	0.5	1232	21	AA273867	Human secreted pro
c 735	0.5	1003	22	AA291893	Human secreted pro	c 808	24	0.5	1235	21	AA252509	Human secreted pro
c 736	0.5	1007	21	AA295508	Human secreted pro	c 809	24	0.5	1236	9	AA280229	Bioadhesive precu
c 737	0.5	1013	12	AA212431	Fragment D of urat	c 810	24	0.5	1236	9	AA282450	Sequence for a gen
c 738	0.5	1016	20	AA297920	Human secreted pro	c 811	24	0.5	1237	21	AA277309	Human ORFX ORF2864
c 739	0.5	1018	20	AA277486	Human ovarian tumo	c 812	24	0.5	1240	21	AA298130	Human colon cancer
c 740	0.5	1020	20	AA252865	Human prostate tum	c 813	24	0.5	1247	21	AA259186	Human secreted pro
c 741	0.5	1020	20	AA281240	Human oligophrenin	c 814	24	0.5	1254	21	AA282973	Cluster-A cDNA enc
											AA295463	Human secreted pro



c 523	24	0.5	566	21	AAZ80382	Human colon cancer	c 596	24	0.5	698	21	AAZ52541	Human secreted pro
c 524	24	0.5	567	20	AAZ51738	DNA encoding a hum	c 597	24	0.5	704	20	AAZ19425	Nicotiana tabacum
c 525	24	0.5	569	21	AAZ97052	Human secreted pro	c 598	24	0.5	705	21	AAZ21389	Human low adenosin
c 526	24	0.5	571	20	AAZ61364	DNA encoding a hum	c 599	24	0.5	705	21	AAZ35267	Human adenosine re
c 527	24	0.5	572	19	AAV44301	Human secreted pro	600	24	0.5	705	21	AAZ80590	Human colon cancer
c 528	24	0.5	575	20	AAV84619	Human secreted pro	c 601	24	0.5	706	20	AAZ87358	Human p2X3 puriner
c 529	24	0.5	578	21	AAZ6861	Human ORFX ORF2416	c 602	24	0.5	706	21	AAZ64100	Human p2X-3 recept
c 530	24	0.5	578	21	AAZ80673	Human colon cancer	c 603	24	0.5	708	20	AAZ611302	DNA encoding a hum
c 531	24	0.5	580	17	AAZ29031	Endoglucanase (Pam	c 604	24	0.5	710	21	AAZ02121	Human colon cancer
c 532	24	0.5	581	21	AAZ15634	Human prostate can	c 605	24	0.5	712	19	AAV59811	Human secreted pro
c 533	24	0.5	581	21	AAZ16007	Human colon cancer	c 606	24	0.5	719	20	AAZ27374	Human secreted pro
c 534	24	0.5	587	21	AAZ80546	Human secreted pro	c 607	24	0.5	722	21	AAZ44453	Arabidopsis thalia
c 535	24	0.5	588	20	AAZ84464	Human secreted pro	c 608	24	0.5	727	21	AAZ97470	Human prostate can
c 536	24	0.5	591	21	AAZ16401	Human colon cancer	c 609	24	0.5	731	16	AAZ099804	PR-1 like gene PR-
c 537	24	0.5	592	22	AAZ74202	DNA encoding eviro	c 610	24	0.5	732	21	AAZ80613	Human colon cancer
c 538	24	0.5	595	20	AAZ41303	Human normal ovari	c 611	24	0.5	732	22	AAZ44861	Human breast cancer
c 539	24	0.5	599	21	AAZ80063	Human colon cancer	c 612	24	0.5	733	20	AAZ41378	Human normal uteru
c 540	24	0.5	602	20	AAZ41243	Human normal ovari	c 613	24	0.5	735	19	AAZ29153	Mouse zins3 gene.
c 541	24	0.5	602	20	AAZ20419	Human secreted pro	c 614	24	0.5	735	21	AAZ50153	Mouse insulin faml
c 542	24	0.5	603	18	AAZ91169	Secreted cytokine	c 615	24	0.5	738	20	AAZ30396	DNA encoding a hum
c 543	24	0.5	603	21	AAZ75983	Human ORFX ORF158	c 616	24	0.5	738	21	AAZ16171	Human prostate can
c 544	24	0.5	604	21	AAZ80307	Human colon cancer	c 617	24	0.5	738	22	AAZ54313	DNA encoding prote
c 545	24	0.5	606	20	AAZ24555	Human lung tumor a	c 618	24	0.5	739	21	AAZ58615	Human PRO1283 prot
c 546	24	0.5	606	21	AAZ65794	Human lung cancer-	c 619	24	0.5	739	21	AAZ37071	Human PRO1283 (UNQ
c 547	24	0.5	607	21	AAZ16283	Human colon cancer	c 620	24	0.5	741	21	AAZ97469	Human prostate can
c 548	24	0.5	608	20	AAZ36688	Mouse OHP106 proti	c 621	24	0.5	745	17	AAZ47800	Oat-derived thioni
c 549	24	0.5	608	21	AAZ16178	Human colon cancer	c 622	24	0.5	745	19	AAZ82783	Clone dn809_5 isol
c 550	24	0.5	608	21	AAZ16518	Human colon cancer	c 623	24	0.5	745	21	AAZ87692	Human secreted pro
c 551	24	0.5	611	21	AAZ16173	Human colon cancer	c 624	24	0.5	749	22	AAZ44900	Human breast cancer
c 552	24	0.5	612	21	AAZ16275	Human colon cancer	c 625	24	0.5	750	21	AAZ59779	Human secreted pro
c 553	24	0.5	620	20	AAZ06238	Human secreted pro	c 626	24	0.5	755	11	AAZ03970	Clone p25b. Mus m
c 554	24	0.5	622	21	AAZ16179	Human colon cancer	c 627	24	0.5	756	11	AAZ039716	Human cancer assoc
c 555	24	0.5	622	21	AAZ16219	Human colon cancer	c 628	24	0.5	767	16	AAZ85889	Wild-type bovine p
c 556	24	0.5	627	21	AAZ16354	Human colon cancer	c 629	24	0.5	769	21	AAZ82122	Arabidopsis thalia
c 557	24	0.5	627	21	AAZ15999	Human colon cancer	c 630	24	0.5	772	22	AAZ32766	Human secreted pro
c 558	24	0.5	628	21	AAZ69661	Human ERK activato	c 631	24	0.5	774	17	AAZ10120	Vascular endotheli
c 559	24	0.5	629	21	AAZ16300	Human colon cancer	c 632	24	0.5	774	18	AAZ85644	Antisense inhibito
c 560	24	0.5	630	20	AAZ36690	Mouse OHP106 prote	c 633	24	0.5	774	19	AAZ79139	Human vascular end
c 561	24	0.5	631	19	AAZ82801	ATG-1120 (allograf	c 634	24	0.5	774	19	AAZ15102	Human vascular end
c 562	24	0.5	631	21	AAZ16349	Human colon cancer	c 635	24	0.5	774	19	AAZ95393	Human secreted pro
c 563	24	0.5	632	21	AAZ16494	Human colon cancer	c 636	24	0.5	781	21	AAZ33349	Human prostate can
c 564	24	0.5	632	21	AAZ80714	Human colon cancer	c 637	24	0.5	784	21	AAZ97286	Human secreted pro
c 565	24	0.5	633	19	AAZ40479	Human secreted pro	c 638	24	0.5	785	20	AAZ24903	Human secreted pro
c 566	24	0.5	638	21	AAZ60768	Pig SENS ligand nu	c 639	24	0.5	787	22	AAZ33240	Human secreted pro
c 567	24	0.5	639	21	AAZ95280	Corn cyclin-depend	c 640	24	0.5	788	11	AAZ001787	Partial sequence o
c 568	24	0.5	639	21	AAZ98055	Human secreted pro	c 641	24	0.5	790	19	AAZ13117	Inserted DNA fragm
c 569	24	0.5	640	20	AAZ30333	DNA encoding a hum	c 642	24	0.5	790	21	AAZ39903	DNA fragment from
c 570	24	0.5	643	20	AAZ84539	Human secreted pro	c 643	24	0.5	797	18	AAZ02898	Human HMGI-C aberr
c 571	24	0.5	643	21	AAZ68086	Human secreted pro	c 644	24	0.5	801	21	AAZ62995	Human ORFX ORF2550
c 572	24	0.5	644	21	AAZ57403	Hepatitis C virus	c 645	24	0.5	805	20	AAZ52972	Human prostate tum
c 573	24	0.5	645	21	AAZ80022	Human colon cancer	c 646	24	0.5	809	21	AAZ68125	Human secreted pro
c 574	24	0.5	649	21	AAZ58893	Human tumour suppr	c 647	24	0.5	813	21	AAZ87676	Human secreted pro
c 575	24	0.5	656	20	AAZ41245	Human normal ovari	c 648	24	0.5	818	20	AAZ15208	Human gene express
c 576	24	0.5	657	21	AAZ69500	Human secreted pro	c 649	24	0.5	821	21	AAZ02473	Human colon cancer
c 577	24	0.5	657	21	AAZ16086	Human colon cancer	c 650	24	0.5	822	21	AAZ59202	Human secreted pro
c 578	24	0.5	658	21	AAZ80732	Human colon cancer	c 651	24	0.5	823	21	AAZ23461	CDNA encoding huma
c 579	24	0.5	660	19	AAZ27070	Double-stranded ad	c 652	24	0.5	826	17	AAZ15229	Tumor necrosis fac
c 580	24	0.5	664	21	AAZ26336	Human secreted pro	c 653	24	0.5	826	18	AAZ94632	TNF-RI-DD ligand p
c 581	24	0.5	667	20	AAZ09052	Differentiation in	c 654	24	0.5	827	21	AAZ78094	Human cancer assoc
c 582	24	0.5	667	21	AAZ46304	Human differentiat	c 655	24	0.5	830	18	AAZ62455	Lambda 5h-1 clone
c 583	24	0.5	668	20	AAZ06772	Marmoset sperm-spe	c 656	24	0.5	831	20	AAZ89609	Human secreted pro
c 584	24	0.5	669	21	AAZ80698	Human colon cancer	c 657	24	0.5	831	20	AAZ37428	Human secreted pro
c 585	24	0.5	685	20	AAZ27366	Human secreted pro	c 658	24	0.5	832	20	AAZ30363	DNA encoding a hum
c 586	24	0.5	685	21	AAZ64992	Membrane-bound pro	c 659	24	0.5	837	21	AAZ69637	Human secreted pro
c 587	24	0.5	685	22	AAZ44138	Human PRO820 (UNQ5	c 660	24	0.5	839	22	AAZ86484	Maize Mac20 coding
c 588	24	0.5	686	20	AAZ61374	DNA encoding a hum	c 661	24	0.5	841	21	AAZ52082	Pigment protein fr
c 589	24	0.5	687	20	AAZ30410	DNA encoding a hum	c 662	24	0.5	843	20	AAZ06774	Marmoset sperm-spe
c 590	24	0.5	688	17	AAZ16856	Integrin subunit b	c 663	24	0.5	846	22	AAZ85089	Atherosclerosis-as
c 591	24	0.5	690	21	AAZ79990	Human colon cancer	c 664	24	0.5	849	21	AAZ59394	Human secreted pro
c 592	24	0.5	693	21	AAZ77669	Human PRO1286 cDNA	c 665	24	0.5	855	21	AAZ21167	Human low adenosin
c 593	24	0.5	693	21	AAZ37065	Human PRO1286 (UNQ	c 666	24	0.5	855	21	AAZ35045	Human adenosine re
c 594	24	0.5	693	22	AAZ54301	DNA encoding prote	c 667	24	0.5	859	20	AAZ97989	Human secreted pro
c 595	24	0.5	693	22	AAZ91565	Human PRO1286 cDNA	c 668	24	0.5	863	18	AAZ91302	Human H1075-1 secr



377	24	0.5	98	21	AAC21458	Human secreted pro	c 450	24	0.5	327	21	AAA31562	Plant microsatelli
378	24	0.5	98	21	AAC21773	Human secreted pro	451	24	0.5	327	21	AAA16539	Human colon cancer
379	24	0.5	100	18	AX83229	Breast cancer tumo	c 452	24	0.5	330	20	AAV83960	Bacterial artifici
380	24	0.5	100	19	AAV68827	DNA molecule encod	c 453	24	0.5	334	21	AAC19789	Human secreted pro
381	24	0.5	100	21	AAC80752	Human breast tumou	c 454	24	0.5	336	21	AAA16503	Human colon cancer
382	24	0.5	101	21	AAC16095	Human secreted pro	c 455	24	0.5	340	21	AAC98349	Human colon cancer
383	24	0.5	102	16	AAQ99624	Homology vector 52	456	24	0.5	342	21	AAA16337	Human colon cancer
384	24	0.5	105	21	AAQ60459	Murine factor V 5'	457	24	0.5	344	22	AAF67203	Novel human polynu
385	24	0.5	105	21	AAC29274	Human secreted pro	458	24	0.5	352	14	AAQ46676	Sequence Inserted
386	24	0.5	106	18	AAV00430	3' fragment of clo	c 459	24	0.5	352	21	AAF16488	Human prostate can
387	24	0.5	106	19	AAV46418	Ribosomal RNA DNA	460	24	0.5	352	21	AAA16524	Human colon cancer
388	24	0.5	108	16	AAQ99635	Homology vector 53	c 461	24	0.5	353	21	AAC77649	Human cancer assoc
389	24	0.5	110	20	AX335561	Secreted protein c	c 462	24	0.5	364	20	AAV71202	Human colon cancer
390	24	0.5	110	21	AAC13607	Human secreted pro	463	24	0.5	364	21	AAV93748	EST clone BN33. H
391	24	0.5	121	21	AAC78441	Human cancer assoc	464	24	0.5	371	22	AAF68094	Cat flea hindgut a
392	24	0.5	130	21	AAC16479	Human secreted pro	c 465	24	0.5	372	21	AAC99865	Human lung tumour
393	24	0.5	140	21	AAC29478	Human secreted pro	c 466	24	0.5	373	21	AAA40500	Human fetal brain
394	24	0.5	160	22	AAF24360	Retroviral recombi	467	24	0.5	373	22	AAF66909	Novel human polynu
395	24	0.5	164	21	AAF18008	Lung cancer associ	468	24	0.5	374	21	AAA16130	Human colon cancer
396	24	0.5	165	17	AAV34610	Probe for detectin	469	24	0.5	378	21	AAAD00146	TRI2 related DNA-1
397	24	0.5	170	21	AAC13889	Human secreted pro	470	24	0.5	378	22	AAF66903	Novel human polynu
398	24	0.5	170	21	AAZ42381	Human 5' EST isola	c 471	24	0.5	386	21	AAC78443	Human cancer assoc
399	24	0.5	178	15	AAQ68946	Junction B of SfII	472	24	0.5	388	22	AAF65550	Novel human polynu
400	24	0.5	178	15	AAQ70553	Junction B of SfII	473	24	0.5	396	22	AAF94836	Human ovarian can
401	24	0.5	178	18	AAV48503	Homology vector 50	474	24	0.5	396	22	AAF64568	Novel human polynu
402	24	0.5	178	20	AAV81149	Sequence of juncti	475	24	0.5	403	21	AAA16024	Human colon cancer
403	24	0.5	178	21	AAC67864	FPV homology vecto	476	24	0.5	404	22	AAF66780	Novel human polynu
404	24	0.5	178	21	AAZ49297	FPV homology vecto	477	24	0.5	406	22	AAF54712	Nucleotide sequenc
405	24	0.5	178	22	AAZ24361	Retroviral recombi	478	24	0.5	406	22	AAF66944	Novel human polynu
406	24	0.5	179	19	AAV11620	Homo sapiens adult	c 479	24	0.5	428	21	AAC55708	Human differential
407	24	0.5	182	15	AAQ70552	Junction A of SfII	c 480	24	0.5	433	18	AAV00428	3' fragment of clo
408	24	0.5	182	18	AAV48502	Homology vector 50	c 481	24	0.5	433	21	AAC98176	Human colon cancer
409	24	0.5	182	20	AAV81148	Sequence of juncti	c 482	24	0.5	433	22	AAC85068	Atherosclerosis-as
410	24	0.5	182	21	AAC67863	FPV homology vecto	c 483	24	0.5	437	18	AAV72938	Phaffia glyceralde
411	24	0.5	182	21	AAC11449	Human secreted pro	484	24	0.5	438	14	AAQ46679	Sequence Inserted
412	24	0.5	182	21	AAZ49296	FPV homology vecto	c 485	24	0.5	441	20	AAZ09474	p135-WT3 construct
413	24	0.5	183	14	AAQ68893	Human brain Expres	c 486	24	0.5	441	22	AAF75092	Human colon associ
414	24	0.5	190	18	AAV73673	Competitor No. 1 u	c 487	24	0.5	450	21	AAA45071	Human secreted exp
415	24	0.5	190	18	AAV73675	Competitor No. 3 u	c 488	24	0.5	460	20	AAV87401	EST clone CD194
416	24	0.5	192	21	AAC13025	Human secreted pro	c 489	24	0.5	466	21	AAC94840	Cat flea hindgut a
417	24	0.5	196	20	AAZ21523	Progression suppre	490	24	0.5	470	19	AAV61312	cDNA sequence of p
418	24	0.5	198	22	AAF60573	Influenza neuramin	491	24	0.5	470	19	AAV58579	Prostate tumour sp
419	24	0.5	205	16	AAQ99629	Homology vector 53	492	24	0.5	470	21	AAA06342	Human immunogenic
420	24	0.5	207	14	AAQ93099	GFP - IE, part of	c 493	24	0.5	477	21	AAC93703	Human secreted pro
421	24	0.5	208	21	AAA42364	Human secreted exp	494	24	0.5	479	21	AAC80348	Human colon cancer
422	24	0.5	208	22	AAV24363	Retroviral recombi	c 495	24	0.5	485	21	AAZ65037	Membrane-bound pro
423	24	0.5	217	21	AAC15434	Human secreted pro	c 496	24	0.5	485	22	AAF92079	Human PRO1003 CDNA
424	24	0.5	218	21	AAA50090	Arabidopsis herbic	c 497	24	0.5	485	22	AAF44183	Human secreted pro
425	24	0.5	220	18	AAV73674	Competitor No. 2 u	c 498	24	0.5	493	21	AAC66424	Human secreted pro
426	24	0.5	224	21	AAA45317	Human secreted exp	499	24	0.5	498	21	AAC93800	Cat flea hindgut a
427	24	0.5	226	21	AAC14683	Human secreted pro	500	24	0.5	498	21	AAC94407	Cat flea hindgut a
428	24	0.5	230	21	AAC29579	Human secreted pro	c 501	24	0.5	506	19	AAV59746	Human secreted pro
429	24	0.5	236	21	AAC93680	Cat flea hindgut a	502	24	0.5	509	19	AAV61342	Extended cDNA seq
430	24	0.5	256	21	AAA16182	Human colon cancer	503	24	0.5	509	19	AAV58672	Prostate tumour sp
431	24	0.5	259	18	AAV91303	Human H1075-1 secr	c 504	24	0.5	509	21	AAC59006	Human secreted pro
432	24	0.5	259	18	AAV00423	3' fragment of clo	505	24	0.5	509	21	AAA08435	Human immunogenic
433	24	0.5	263	18	AAV85891	Malassezia fungus	c 506	24	0.5	517	22	AAA45399	Plant defensin cod
434	24	0.5	263	20	AAV88738	EST clone HJ674	c 507	24	0.5	522	21	AAF12440	Aspergillus oryzae
435	24	0.5	277	19	AAA44945	Human secreted exp	508	24	0.5	523	14	AAQ46677	Sequence Inserted
436	24	0.5	285	19	AAV11623	Homo sapiens adult	509	24	0.5	526	16	AAQ86309	Wilson disease gen
437	24	0.5	288	20	AAV89510	EST clone COL151	c 510	24	0.5	534	21	AAA38006	UCK-1 nucleotide s
438	24	0.5	296	21	AAV16014	Human colon cancer	511	24	0.5	537	21	AAZ57401	Hepatitis C virus
439	24	0.5	299	19	AAV69628	Human secreted pro	c 512	24	0.5	541	20	AAZ87415	Hepatocellular car
440	24	0.5	301	14	AAQ59562	Human brain Expres	c 513	24	0.5	543	21	AAC81743	Human secreted pro
441	24	0.5	304	20	AAZ21531	Progression elevat	c 514	24	0.5	546	22	AAF44946	Human breast cance
442	24	0.5	305	20	AAV86214	EST clone M338. H	c 515	24	0.5	549	19	AAV38759	CDNA nfSP17-549 en
443	24	0.5	307	22	AAF68866	Human lung tumour	516	24	0.5	549	19	AAV38760	Complementary stra
444	24	0.5	318	14	AAQ59992	Human brain Expres	c 517	24	0.5	550	21	AAZ52554	Human secreted pro
445	24	0.5	318	18	AAV91312	Human J143-1 secre	518	24	0.5	553	19	AAV46447	Human chromosome 7
446	24	0.5	322	21	AAV00436	3' fragment of clo	c 519	24	0.5	556	19	AAV66764	Pathogen response
447	24	0.5	322	21	AAC98002	Human colon cancer	c 520	24	0.5	563	21	AAA50065	Prostate cancer as
448	24	0.5	322	21	AAA16211	Human colon cancer	c 521	24	0.5	565	21	AAA04013	Potato infection i
449	24	0.5	322	21	AAA16357	Human colon cancer	c 522	24	0.5	565	21	AAA45987	Human metastatic m



c 231	25	0.5	2730	22	AAE29459	Human TANGO 332 cD	c 304	25	0.5	12827	19	AAV09036	Equine arthritis v
c 232	25	0.5	2738	19	AAV48220	APEG-1 5' vascular	c 305	25	0.5	13104	14	AAQ46852	Clone of recombin
c 233	25	0.5	2738	21	AAV51046	2.7 kb fragment co	c 306	25	0.5	13206	11	AAQ05243	Rat acyl peptide h
c 234	25	0.5	2812	21	AAE16219	Human prostate can	c 307	25	0.5	15528	19	AAV09039	Equine arthritis v
c 235	25	0.5	2830	20	AAE27909	Canine B7-1 protei	c 308	25	0.5	16956	21	AAZ91923	Wild type (C57BL/6
c 236	25	0.5	2830	20	AAE27910	Canine B7-1 gene c	c 309	25	0.5	20303	18	AAT71699	Human deoxycytidyl
c 237	25	0.5	2840	21	AAE78596	Human PRO4993 nucl	c 310	25	0.5	21742	21	AAF20938	Human high affinit
c 238	25	0.5	2980	21	AAZ23444	cDNA encoding huma	c 311	25	0.5	21742	21	AAA34816	Human adenosine re
c 239	25	0.5	3046	19	AAE29074	Rattus norvegicus	c 312	25	0.5	22540	21	AAC88442	Thiamine responsiv
c 240	25	0.5	3080	21	AAE78415	Human secreted pro	c 313	25	0.5	24000	21	AAA88551	Human dual-specific
c 241	25	0.5	3114	18	AAV02800	Human RHAMM cDNA	c 314	25	0.5	25871	21	AAA09888	Human gonadotropin
c 242	25	0.5	3212	22	AAC84373	Nucleotide sequenc	c 315	25	0.5	26016	19	AAV75272	Human flavin-conta
c 243	25	0.5	3355	21	AAE16121	Human prostate can	c 316	25	0.5	26764	18	AAT71696	Human deoxycytidyl
c 244	25	0.5	3413	21	AAE251048	Mouse APEG-1 gene	c 317	25	0.5	33030	22	AAE29337	Atopy related gene
c 245	25	0.5	3451	20	AAE35701	cDNA encoding a pr	c 318	25	0.5	34488	22	AAE97854	Human neuroblastom
c 246	25	0.5	3455	22	AAE32241	Mouse urinary triyp	c 319	25	0.5	35455	22	AAE54723	Nucleotide sequenc
c 247	25	0.5	3657	21	AAC76218	Human ORFX ORF1773	c 320	25	0.5	39796	21	AAC61681	Nucleotide sequenc
c 248	25	0.5	3721	21	AAE96343	cDNA encoding a no	c 321	25	0.5	45546	20	AAE23520	Human kidney amino
c 249	25	0.5	3729	22	AAE77688	Human wild-type Fc	c 322	25	0.5	49999	20	AAE23891	Murine LOBO genom
c 250	25	0.5	3798	14	AAQ42424	ADA2 DNA. Synthet	c 323	25	0.5	49999	20	AAE23896	Murine LOBO homolo
c 251	25	0.5	3839	22	AAE32759	Human secreted pro	c 324	25	0.5	49999	20	AAE23904	Human LOBO homolog
c 252	25	0.5	3842	20	AAE23533	Tomato Xa21 clone	c 325	25	0.5	50000	21	AAA96364	Polymorphic repeat
c 253	25	0.5	3851	21	AAC77558	Human ORFX ORF3113	c 326	25	0.5	72604	20	AAE10752	Genomic sequence o
c 254	25	0.5	3901	21	AAE33361	Human secreted pro	c 327	25	0.5	78925	21	AAC89888	Human FN gene. HO
c 255	25	0.5	3938	20	AAE05823	Rat pheromone rece	c 328	25	0.5	80595	20	AAE83939	HC-contig derived
c 256	25	0.5	4066	19	AAE59181	Human bak gene pro	c 329	25	0.5	81001	22	AAE93035	Human apolipoprote
c 257	25	0.5	4131	22	AAE77689	Human variant Fcpe	c 330	25	0.5	81369	21	AAE97997	Human T gene DNA
c 258	25	0.5	4180	10	AAE91773	Rat androgen recep	c 331	25	0.5	87350	18	AAE83003	Human WRN genomic
c 259	25	0.5	4321	19	AAE52853	Human eyal-c encod	c 332	25	0.5	106746	21	AAE10225	Human PCTA-1 genom
c 260	25	0.5	4366	19	AAE94063	Human growth hormo	c 333	25	0.5	117609	21	AAE21435	Human receptor-rel
c 261	25	0.5	4436	9	AAE81716	Human growth hormo	c 334	25	0.5	121162	21	AAC66548	Human kinasin-like
c 262	25	0.5	4455	19	AAE15448	Arabidopsis pr-1 p	c 335	25	0.5	138169	21	AAA34791	Human adenosine re
c 263	25	0.5	4698	13	AAE32370	PLA genomic DNA	c 336	25	0.5	141589	21	AAE20913	Human ELAM-1 polyn
c 264	25	0.5	4698	15	AAE72475	Entire PLA genomic	c 337	25	0.5	141589	21	AAE21127	Human low adenosin
c 265	25	0.5	4698	16	AAE01176	P815A antigen prec	c 338	25	0.5	141589	21	AAE21152	Human adenosine re
c 266	25	0.5	4698	17	AAE14692	MAGE genomic DNA	c 339	25	0.5	141589	21	AAA35005	Human adenosine re
c 267	25	0.5	4749	18	AAE84111	PLA genomic sequen	c 340	25	0.5	141589	21	AAA35030	Human adenosine re
c 268	25	0.5	5235	21	AAE59230	FSPI promoter link	c 341	25	0.5	146981	21	AAE21442	Human ELAM-1 polyn
c 269	25	0.5	5235	21	AAE76134	Human OREX ORF1689	c 342	25	0.5	151826	21	AAE22291	BAC containing rep
c 270	25	0.5	5261	16	AAE09602	Human Ah receptor	c 343	25	0.5	160552	22	AAE02697	Human glycosyl sul
c 271	25	0.5	5261	18	AAE85436	Human Ah-receptor	c 344	25	0.5	162450	21	AAE86967	Human glycosyl sul
c 272	25	0.5	5301	18	AAE93627	Human Ah-receptor	c 345	25	0.5	162450	21	AAE86967	Human glycosyl sul
c 273	25	0.5	5305	20	AAE28289	Human metastasis-a	c 346	25	0.5	209273	21	AAE21437	Retinoblastoma bin
c 274	25	0.5	5555	15	AAE64898	Rat neuronal immed	c 347	25	0.5	240825	22	AAE24497	Human factor-relat
c 275	25	0.5	5555	17	AAE18533	HER4 with alternat	c 348	24	0.5	38	21	AAE243893	Human PC-3 gene
c 276	25	0.5	5607	21	AAE88884	Receptor tyrosine	c 349	24	0.5	40	13	AAE25027	M. tuberculosis ip
c 277	25	0.5	5607	21	AAE47788	Human Wnt-1 gene	c 350	24	0.5	40	13	AAE25027	Anti-sense oligonu
c 278	25	0.5	5836	16	AAE96296	Human Wnt-1 encodi	c 351	24	0.5	40	15	AAE55168	Sequence of primer
c 279	25	0.5	5836	18	AAE84444	Human prostate-sp	c 352	24	0.5	47	20	AAE201041	Probe for human PG
c 280	25	0.5	5836	21	AAE60485	Human prostate-spe	c 353	24	0.5	48	16	AAE04081	Trypsin inhibitory
c 281	25	0.5	5836	21	AAE60485	5' flanking region	c 354	24	0.5	48	16	AAE04081	Primer S1N3ball700
c 282	25	0.5	5836	21	AAE60485	Nucleotide sequenc	c 355	24	0.5	48	17	AAE35073	Sindbis-based tum
c 283	25	0.5	5836	21	AAE39130	Human prostate spe	c 356	24	0.5	48	17	AAE35073	Sindbis PCR primer
c 284	25	0.5	5888	20	AAE299332	DNA sequence of co	c 357	24	0.5	48	19	AAE42384	Reverse PCR primer
c 285	25	0.5	6083	21	AAE51398	Human GDNF fragmen	c 358	24	0.5	48	20	AAE70704	Reverse PCR primer
c 286	25	0.5	6792	21	AAE294062	Chromosome 16g tum	c 359	24	0.5	48	21	AAE292785	Reverse PCR primer
c 287	25	0.5	7063	20	AAE81916	Aryl hydrocarbon r	c 360	24	0.5	48	21	AAE292785	Sindbis basic vect
c 288	25	0.5	7063	20	AAE81916	cDNA encoding huma	c 361	24	0.5	53	21	AAE18812	Sindbis basic vect
c 289	25	0.5	7397	19	AAE60578	cDNA encoding huma	c 362	24	0.5	54	21	AAE18812	Human secreted pro
c 290	25	0.5	7452	16	AAE85594	Human tumour suppr	c 363	24	0.5	56	17	AAE17032	Human secreted pro
c 291	25	0.5	7452	16	AAE85594	Human glycoprotein	c 364	24	0.5	60	18	AAE69401	Human secreted pro
c 292	25	0.5	7452	21	AAE61217	DNA encoding huma	c 365	24	0.5	60	21	AAE14231	Human secreted pro
c 293	25	0.5	7452	22	AAE90646	Human secreted pro	c 366	24	0.5	60	22	AAE62064	Human secreted pro
c 294	25	0.5	8365	20	AAE20056	Plasmodium falcipa	c 367	24	0.5	67	21	AAE14340	Human secreted pro
c 295	25	0.5	10120	20	AAE233683	Human DKC1 DNA fra	c 368	24	0.5	69	21	AAE12637	Human secreted pro
c 296	25	0.5	10120	20	AAE233683	Human DKC1 DNA fra	c 369	24	0.5	72	21	AAE27585	Human secreted pro
c 297	25	0.5	11298	18	AAE86756	Human high affinit	c 370	24	0.5	73	21	AAE12555	Human secreted pro
c 298	25	0.5	11298	19	AAE54661	Human beta subunit	c 371	24	0.5	77	21	AAE14910	Human secreted pro
c 299	25	0.5	11298	21	AAE20937	Human high affinit	c 372	24	0.5	83	21	AAE13898	Human secreted pro
c 300	25	0.5	11298	22	AAE34815	Human adenosine re	c 373	24	0.5	85	19	AAE32418	Homo sapiens clone
c 301	25	0.5	11357	14	AAE92144	Human IGERB gene S	c 374	24	0.5	87	21	AAE23541	Human secreted pro
c 302	25	0.5	11967	22	AAE97863	Human FCER1 beta c	c 375	24	0.5	90	21	AAE22118	Human secreted pro
c 303	25	0.5	12117	21	AAE96368	Human neuroblastom	c 376	24	0.5	92	18	AAE88439	Padlock probe alph
						Polymorphic repeat				98	20	AAE00174	Porcine reproducti



c 85	25	0.5	294	20	AAV82725	Partial sequence o	c 158	25	0.5	1149	21	AAA58565	DNA encoding human
c 86	25	0.5	294	21	AAA89542	Xyloglucan endoglu	c 159	25	0.5	1173	21	AAA63941	Nucleotide sequenc
c 87	25	0.5	294	21	AAA90868	Xyloglucan endoglu	c 160	25	0.5	1192	21	AAA51406	Chromosome 16q tum
c 88	25	0.5	294	22	AAAF77876	Partial XGU endogl	c 161	25	0.5	1197	17	AAAT32429	Wasp venom BrhTX-1
c 89	25	0.5	294	22	AAAF79737	Aspergillus aculea	c 162	25	0.5	1197	19	AAV17146	Insecticidal toxin
c 90	25	0.5	294	22	AAAF81169	Aspergillus aculea	c 163	25	0.5	1201	17	AAAT32430	Wasp venom BrhTX-1
c 91	25	0.5	294	22	AAAF75790	Partial XGU endogl	c 164	25	0.5	1210	20	AAZ50064	Impatiens balsamia
c 92	25	0.5	294	22	AAAF77567	A aculeatus xylogl	c 165	25	0.5	1240	20	AAZ41296	Human normal ovary
c 93	25	0.5	294	22	AAAF57479	3' portion of cDNA	c 166	25	0.5	1259	19	AAV52846	Human eyal gene co
c 94	25	0.5	327	18	AAAT8091	Human secreted pro	c 167	25	0.5	1284	21	AAV59950	Human secreted pro
c 95	25	0.5	327	18	AAV02166	Human secreted pro	c 168	25	0.5	1285	20	AAV84620	Human TNF11 coding
c 96	25	0.5	387	22	AAV65820	Novel human polynu	c 169	25	0.5	1285	20	AAV84620	Human TNF11 coding
c 97	25	0.5	390	21	AAA16227	Human colon cancer	c 170	25	0.5	1302	12	AAQ11112	MHC class II anti
c 98	25	0.5	416	21	AAAC99917	Human secreted pro	c 171	25	0.5	1321	21	AAV35361	Arabidopsis thalia
c 99	25	0.5	419	21	AAAC74314	Human secreted pro	c 172	25	0.5	1322	20	AAV04402	Human secreted pro
c 100	25	0.5	434	21	AAAC74424	Human secreted pro	c 173	25	0.5	1324	20	AAV04402	Human secreted pro
c 101	25	0.5	443	21	AAV75749	cDNA encoding a de	c 174	25	0.5	1337	20	AAV04345	Human secreted pro
c 102	25	0.5	484	22	AAAF28388	Anopheles albimanu	c 175	25	0.5	1348	19	AAV34289	Human secreted pro
c 103	25	0.5	484	22	AAAF28388	Anopheles albimanu	c 176	25	0.5	1353	21	AAV18255	Lung cancer associ
c 104	25	0.5	499	21	AAZ97350	Barley Clone F. H	c 177	25	0.5	1389	21	AAV79822	Human secreted pro
c 105	25	0.5	500	21	AAAC93907	Human prostate can	c 178	25	0.5	1396	16	AAQ98498	Human secreted pro
c 106	25	0.5	528	18	AAV02891	Cat flea hindgut a	c 179	25	0.5	1396	18	AAV97324	Human (50 kDa) dylr
c 107	25	0.5	528	18	AAV02891	Human HMGI-C aberr	c 180	25	0.5	1442	21	AAV80566	Human secreted pro
c 108	25	0.5	551	21	AAV21818	Human secreted pro	c 181	25	0.5	1444	21	AAV77169	Human ORFX ORF2724
c 109	25	0.5	553	19	AAV34276	Human breast and o	c 182	25	0.5	1447	13	AAQ30002	Human ORFX ORF2724
c 110	25	0.5	554	22	AAAF33138	Human secreted pro	c 183	25	0.5	1447	15	AAV58686	HCNP precursor gen
c 111	25	0.5	557	19	AAV34223	Human secreted pro	c 184	25	0.5	1467	21	AAZ52508	Human hippocampal
c 112	25	0.5	569	20	AAV27389	Human secreted pro	c 185	25	0.5	1487	20	AAV00715	Human secreted pro
c 113	25	0.5	571	20	AAZ06228	Human secreted pro	c 186	25	0.5	1487	21	AAV297117	Human secreted pro
c 114	25	0.5	593	21	AAV26329	Human secreted pro	c 187	25	0.5	1492	21	AAV51403	Chromosome 16q tum
c 115	25	0.5	604	21	AAV59787	Human secreted pro	c 188	25	0.5	1494	11	AAQ03369	Sequence encoding
c 116	25	0.5	611	21	AAV16388	Human colon cancer	c 189	25	0.5	1496	10	AAV92386	DNA coding for the
c 117	25	0.5	615	21	AAV16139	Human colon cancer	c 190	25	0.5	1512	18	AAV21272	Alzheimer's diseas
c 118	25	0.5	615	18	AAV02881	Human HMGI-C aberr	c 191	25	0.5	1534	22	AAV23910	Human secreted pro
c 119	25	0.5	617	18	AAV02870	Human HMGI-C/PSAFP	c 192	25	0.5	1558	20	AAV84453	Human secreted pro
c 120	25	0.5	618	21	AAV03065	Arabidopsis thalia	c 193	25	0.5	1563	20	AAV53007	Human prostate tum
c 121	25	0.5	625	21	AAV32915	Arabidopsis thalia	c 194	25	0.5	1570	21	AAV37103	Human secreted pro
c 122	25	0.5	636	21	AAV59368	Human secreted pro	c 195	25	0.5	1570	22	AAV54409	Human secreted pro
c 123	25	0.5	653	19	AAV52938	Pig F2CD antigen e	c 196	25	0.5	1622	21	AAV59771	Human breast and o
c 124	25	0.5	658	21	AAV16322	Human colon cancer	c 197	25	0.5	1645	21	AAV21831	Human breast and o
c 125	25	0.5	689	18	AAV02893	Human HMGI-C aberr	c 198	25	0.5	1660	19	AAV48147	Nicotianamine amin
c 126	25	0.5	707	22	AAV27958	Human dual specifi	c 199	25	0.5	1671	22	AAV33038	Human secreted pro
c 127	25	0.5	715	20	AAV37514	Human secreted pro	c 200	25	0.5	1682	21	AAV18279	Lung cancer associ
c 128	25	0.5	725	21	AAV77602	Human ORFX ORF3157	c 201	25	0.5	1693	22	AAV63803	Human secreted pro
c 129	25	0.5	726	20	AAV03016	Human IL-1ra BAC c	c 202	25	0.5	1831	21	AAV79902	Human secreted pro
c 130	25	0.5	731	13	AAQ31693	RsaI restriction f	c 203	25	0.5	1839	21	AAV78067	Human cancer assoc
c 131	25	0.5	732	21	AAV52519	Human secreted pro	c 204	25	0.5	1894	20	AAV24072	Human hRNP G DNA
c 132	25	0.5	740	20	AAV52980	Human prostate tum	c 205	25	0.5	1927	21	AAV79919	Human secreted pro
c 133	25	0.5	757	21	AAV79355	Plant viral move	c 206	25	0.5	1955	21	AAV93102	Human secreted pro
c 134	25	0.5	815	20	AAV52884	Human prostate tum	c 207	25	0.5	1986	21	AAV95489	Human secreted pro
c 135	25	0.5	825	18	AAV80699	Partial cDNA clone	c 208	25	0.5	2020	22	AAV02199	Chondrus crispus s
c 136	25	0.5	825	18	AAV02157	Human secreted pro	c 209	25	0.5	2058	14	AAQ41227	Clone GP46 encodin
c 137	25	0.5	825	21	AAV040502	Human adult blood	c 210	25	0.5	2085	21	AAV93408	Human secreted pro
c 138	25	0.5	860	21	AAV79359	Plant viral move	c 211	25	0.5	2094	20	AAV19489	Human secreted pro
c 139	25	0.5	870	21	AAV99870	Human secreted pro	c 212	25	0.5	2094	20	AAV99870	Human secreted pro
c 140	25	0.5	876	21	AAV63439	Human secreted pro	c 213	25	0.5	2114	20	AAV19958	Rat TANGO-76 encod
c 141	25	0.5	880	21	AAV21883	Human breast and o	c 214	25	0.5	2139	21	AAV79860	Human secreted pro
c 142	25	0.5	880	21	AAV66182	Human trypsin hL g	c 215	25	0.5	2201	21	AAZ43777	Human fetal brain
c 143	25	0.5	890	21	AAV292417	cDNA encoding huma	c 216	25	0.5	2202	21	AAV52286	Maize replication
c 144	25	0.5	903	21	AAV64431	DNA encoding a hum	c 217	25	0.5	2226	15	AAV072270	Rat MR22 serotonin
c 145	25	0.5	911	21	AAV59124	Human secreted pro	c 218	25	0.5	2238	21	AAV59408	Human secreted pro
c 146	25	0.5	971	21	AAV58883	Human tumour suppr	c 219	25	0.5	2270	20	AAV24865	Human secreted pro
c 147	25	0.5	979	19	AAV69626	Human secreted pro	c 220	25	0.5	2328	18	AAV95227	Human islet cell a
c 148	25	0.5	989	20	AAV33618	Human breast tumou	c 221	25	0.5	2387	21	AAV22374	Human secreted pro
c 149	25	0.5	1013	21	AAV55487	Canine Flt-3 ligand	c 222	25	0.5	2394	22	AAV58095	Atherosclerosis-as
c 150	25	0.5	1013	21	AAV55488	Canine Flt-3 ligand	c 223	25	0.5	2402	19	AAV07202	Mouse epandymin-11
c 151	25	0.5	1024	21	AAV297148	Human prostate can	c 224	25	0.5	2432	20	AAV00650	Human secreted pro
c 152	25	0.5	1027	22	AAV89255	Mouse camello 1 (M	c 225	25	0.5	2435	20	AAV00707	Human secreted pro
c 153	25	0.5	1040	18	AAV02818	Human RRAM genom	c 226	25	0.5	2438	22	AAV92529	Rat T2R02 nucleoti
c 154	25	0.5	1054	21	AAV74245	Human secreted pro	c 227	25	0.5	2574	21	AAV27982	Human homology to
c 155	25	0.5	1076	21	AAV77715	Human cancer assoc	c 228	25	0.5	2599	21	AAV76501	Human ORFX ORF2056
c 156	25	0.5	1130	21	AAV39064	Human secreted pro	c 229	25	0.5	2625	21	AAV16658	Human secreted pro
c 157	25	0.5	1138	21	AAV58267	Rice cellulose syn	c 230	25	0.5	2628	18	AAV01060	Human PKR gene. H



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 16:19:42 ; Search time 299.68 Seconds  
(without alignments)  
11044.012 Million cell updates/sec

Title: US-09-512-581-1

Perfect score: 5271

Sequence: 1 ccggagagcccgagtgag.....gtaaaaaaaaaaaaaaaa 5271

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N\_Geneseq\_0601.\*

1:	/SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5271	100.0	5271 21	Human androgen shu
2	5271	100.0	5355 21	Human androgen shu
3	407	7.7	439 20	EST clone DY17. H
4	348	6.6	738 20	Human gene express
5	292	5.5	295 21	Human secreted pro
6	268	5.1	351 21	Human 5' EST isola
7	219	4.2	284 20	Human gene express
8	131	2.5	405 16	Human gene express
9	120	2.3	161 21	Human gene signatu
10	99	1.9	150 19	Human secreted pro
11	99	1.9	150 19	Human biallelic po

c	12	97	1.8	530	21	AAZ80598	Human colon cancer
c	13	29	0.6	421	20	AAZ21198	Polynucleotide seq
c	14	29	0.6	489	20	AAV87322	EST clone BP810.
c	15	29	0.6	14784	21	AAA64141	Nucleotide sequenc
c	16	29	0.6	45265	21	AAZ46508	Sequence of a COSM
c	17	28	0.5	60	21	AAC23080	Human secreted pro
c	18	28	0.5	1057	22	AAZ02007	Human TAC1-ligand
c	19	27	0.5	680	21	AAC74320	Human secreted pro
c	20	27	0.5	704	20	AAZ40124	Gastric cancer ass
c	21	27	0.5	845	21	AAZ59318	Human secreted pro
c	22	27	0.5	1352	20	AAZ98002	Human secreted pro
c	23	27	0.5	1358	21	AAC59286	Human secreted pro
c	24	27	0.5	1419	21	AAC59246	Human secreted pro
c	25	27	0.5	1776	20	AAC33581	Human breast tumou
c	26	27	0.5	2477	20	AAC34387	Human PRO1082 nucl
c	27	27	0.5	2477	21	AAC78579	Human PRO1082 (UMG
c	28	27	0.5	3076	20	AAZ02987	Human IL-1ra BAC c
c	29	27	0.5	3447	19	AAV16444	Glucanase II gene
c	30	27	0.5	28690	22	AAZ57718	Human IL4Ralpha qe
c	31	27	0.5	43069	21	AAZ36335	Genomic sequence o
c	32	26	0.5	128	21	AAZ00579	Human Hscd66 intro
c	33	26	0.5	182	15	AAZ68945	Junction A of Sf11
c	34	26	0.5	293	20	AAV89475	EST clone CO209.
c	35	26	0.5	394	18	AAZ8088	3' portion of cDNA
c	36	26	0.5	394	18	AAZ02161	Human secreted pro
c	37	26	0.5	451	21	AAC93441	Human secreted pro
c	38	26	0.5	495	18	AAZ02892	Human HMGI-C aberr
c	39	26	0.5	535	21	AAA16380	Human colon cancer
c	40	26	0.5	635	16	AAZ75445	Murine glycosylati
c	41	26	0.5	635	17	AAZ38390	Murine glycosylati
c	42	26	0.5	723	18	AAZ42755	Great tit CHD-W qe
c	43	26	0.5	794	21	AAZ79751	Human secreted pro
c	44	26	0.5	807	17	AAZ27152	Human Machado-Jose
c	45	26	0.5	1067	21	AAZ52545	Human secreted pro
c	46	26	0.5	1216	21	AAZ18145	Lung cancer associ
c	47	26	0.5	1346	21	AAZ79040	Human secreted pro
c	48	26	0.5	1423	10	AAZ90249	Cucumber mosaic v1
c	49	26	0.5	1426	18	AAZ72272	Cucumber mosaic v1
c	50	26	0.5	1459	21	AAZ27155	Human P2 DNA. Hom
c	51	26	0.5	1504	21	AAZ93333	Human secreted pro
c	52	26	0.5	1573	22	AAZ72752	Human prostate can
c	53	26	0.5	1657	21	AAZ27156	Exemplary human P2
c	54	26	0.5	1776	20	AAZ23425	Human Machado-Jose
c	55	26	0.5	2318	21	AAZ59907	Human secreted pro
c	56	26	0.5	2470	21	AAZ20859	Human adenosine A3
c	57	26	0.5	2470	21	AAZ20870	Human adenosine A3
c	58	26	0.5	2470	21	AAZ34737	Human adenosine re
c	59	26	0.5	2470	21	AAZ34748	Human adenosine re
c	60	26	0.5	2588	21	AAZ76756	Human ORFX ORF2311
c	61	26	0.5	2923	20	AAZ04326	Human secreted pro
c	62	26	0.5	3350	17	AAZ08553	Oncogene R-ras mut
c	63	26	0.5	3744	21	AAZ27157	Human P2 genomic D
c	64	26	0.5	5318	22	AAZ59961	Human eukaryotic 1
c	65	26	0.5	11395	21	AAZ21440	Human adenosine A3
c	66	26	0.5	99999	20	AAZ23903	Human LOBO homolog
c	67	26	0.5	72604	20	AAZ10752	Genomic sequence o
c	68	26	0.5	90050	21	AAZ91925	Wild type (C57BL/6
c	69	26	0.5	117609	21	AAZ21435	Human receptor-rel
c	70	26	0.5	121162	21	AAZ66548	Human kinesin-like
c	71	25	0.5	48	20	AAZ57022	WO9923258 oligonuc
c	72	25	0.5	54	15	AAZ063929	Synthetic Cholera
c	73	25	0.5	55	18	AAZ72962	Omega probe P302 u
c	74	25	0.5	75	17	AAZ17036	Human mitochondria
c	75	25	0.5	121	19	AAZ54078	Nucleotide sequenc
c	76	25	0.5	164	21	AAC16464	Human secreted pro
c	77	25	0.5	186	21	AAC28130	Human secreted pro
c	78	25	0.5	226	17	AAZ43285	Promoter for infec
c	79	25	0.5	232	21	AAC20507	Human secreted pro
c	80	25	0.5	254	21	AAC11615	Human secreted pro
c	81	25	0.5	265	21	AAA42328	Human secreted exp
c	82	25	0.5	283	20	AAZ89992	EST clone CW762.
c	83	25	0.5	284	20	AAZ90426	EST clone DM360.
c	84	25	0.5	294	15	AAZ067025	Endoglucanase type







## AUTHORS

Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,  
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,  
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang  
Y., Gu,Y., Chen,Z., and Han,Z.  
Homo sapiens cDNA GK- clones  
Unpublished (2000)

## TITLE

## JOURNAL

## COMMENT

Contact: Zequang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

## FEATURES

## source

1..608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GKATF08"  
/clone\_lib="GK"  
/tissue\_type="hepatocellular carcinoma"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

## BASE COUNT

181 a 113 c 125 g 189 t

## ORIGIN

## Query Match

Best Local Similarity 2.9%; Score 151; DB 32; Length 608;

Matches 351; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 528 accagctatcacagaaccttatttttcagttataaacaatggccacaatcagaagtcacat 587

|||||

Db 80 ACCAGCTATACAGAACCCTTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCAT 139

|||||

QY 588 atgcacatggtagaccttactagctctattttgtgaagtgatcacagtgctccaggag 647

|||||

Db 140 ATGCACATGGTAGACCTTATGAGCTCTATTATTGTGAAGGTGATACAGTGCTCAGGAG 199

|||||

QY 648 ctttggatacgttttagtaaatctggtacctgctcataagaatttaacaagaagca 707

|||||

Db 200 CTTTGGATACGGATTAGTAATCTGCTACCTGCTCATTAAGAATTTAAGCAAGCAAGCA 259

|||||

QY 708 tatgattggcaaggcttactgaagagagacagctcaagctattgagccatatatacc 767

|||||

Db 260 TATGATTTGGCAAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATTATACC 319

|||||

QY 768 acttttttaatcaggtctctgatgtctgggaaaaacatctatcagcgatttgcagagcat 827

|||||

Db 320 AATTTTTTATCAGGATCTGATGCTTTGGGAAAACATCTATCAGCGATTGTCAGACAT 379

|||||

QY 828 gtcttggacttaatttggagctcacaaattatgatagtcatttgcgtctctcg 882

|||||

Db 380 GTCCTTGACTTAATTTGGAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTG 434

|||||

## RESULT 50

## AL121425

## LOCUS

AL121425 167 bp mRNA EST 25-FEB-2000

DEFINITION DKFZp762H077.r1.762 (synonym: hmcl2) Homo sapiens cDNA clone

sequence.

## ACCESSION

AL121425

## VERSION

AL121425.1 GI:5927426

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and  
Wiemann,S.

## TITLE

## JOURNAL

## COMMENT

EST (Ottenwaelder, et al.)  
Unpublished (1999)  
Contact: Ottenwaelder B  
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by Medigenomix (Martinsried/Germany) within the cDNA

sequencing consortium of the German Genome Project. No sl sequence

available.

This clone (DKFZp762H077) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

## source

1..167  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp762H077"  
/clone\_lib="762 (synonym: hmcl2)"  
/tissue\_type="melanoma (MeWo cell line)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"  
BASE COUNT 72 a 14 c 19 g 62 t

## ORIGIN

## Query Match

Best Local Similarity 2.8%; Score 147; DB 105; Length 167;

Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5102 tggcaatgaaattttaagaagaagatttaagatttttaatttttaagagtggttat 5161

|||||

Db 1 TGGCAATGAAATTTTAAAGAAAGATTAAAGTATTTTAAAGAGTGTGTAT 60

|||||

QY 5162 aaaataatgtactgaattctttatccccattttatccatcttccagtttttattaatctac 5221

|||||

Db 61 AAATAATGTAATCTTTTATCCCATTTTATCATCTTCAGTTTTTATTATTAATCTAC 120

|||||

QY 5222 tgatcaataaaattctgtaattgaa 5248

|||||

Db 121 TGTATCAATAAAATTCGTAATTGAA 147

Search completed: September 25, 2001, 20:54:46

Job time: 22799 sec



```

RESULT 47
F12354
LOCUS
DEFINITION HSC38H111 normalized infant brain cDNA Homo sapiens cDNA clone
C-38h11, mRNA sequence.
ACCESSION F12354
VERSION F12354.1
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-38h11
Seq primer: (-21)M13-universal.
FEATURES
source
1..316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-38h11"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/notes="Organ: brain; Vector: lafmid BA; Site.1: HindIII;
Site.2: NotI; sex:Female; dev_stage=3 months old;
Isolate=muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 121 a 69 c 76 g 50 t
ORIGIN
Query Match 3.1%; Score 164; DB 157; Length 316;
Best Local Similarity 100.0%; Pred. No. 6.7e-61;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4074 aggaacagaagtccaaagcaacagcaccgagtgctcaggagagcagcagagagca 4133
|||||
Db 1 ACGGAACAGAAGTCCAAAAGCAACAGCACCAGTCTCAAGGAGAGCACAGCAGAGAGCA 60
QY 4134 gaatctcctgaatctagtgcattgaatccacacagtcacacacagaaagcagga 4193
|||||
Db 61 GAATCTCCCTGAATCTAGTGAATGAATCCACACAGTCCACACAGAAAGGACGAGGA 120
QY 4194 agaccatcaaaacgcattcaccattcacacaaacaaataatgt 4237
|||||
Db 121 AGACCATCAAAACGCCATCACCATCACCACCAACCAAAAAAATGT 164
RESULT 48
BG546853
LOCUS
DEFINITION 602574035F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702089 5',
mRNA EST
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 608)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1538 row: f column: 10
High quality sequence stop: 606.
FEATURES
source
1..606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4702089"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site.1:
SfiI (ggcgctcgcc); Site.2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 190 a 91 c 139 g 186 t
ORIGIN
Query Match 2.9%; Score 152; DB 155; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.2e-55;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 ggagtagcgagtcgcgaacccggagggttagaataatttctgctcagtcattcaaga 81
|||||
Db 3 GGAGTAGCGAGTCGCGAACCGGAGGGGTAGAAATATTTCTGTCATGGCTCATTAAGA 62
QY 82 ctaggaccaatgatgaaaaattacatattccgctgggtcaagaaataatcagataaaa 141
|||||
Db 63 CTAGGACCACATGATGGAATAATACATATCCGCTGGGTCAAGGAATAATCATGATAAA 122
QY 142 tatctaaagaggagatggtgagcagcattaaag 173
|||||
Db 123 TATCTAAAGAGGAGATGGTGAGACGATTAAAG 154
RESULT 49
AV685306
LOCUS
DEFINITION AV685306 GKC Homo sapiens cDNA clone GKCATF08 5', mRNA sequence.
ACCESSION AV685306
VERSION AV685306.1
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 608)

```



```

Db      6 TGA 4

RESULT 45
AA687846
LOCUS   AA687846
DEFINITION nv09a11.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1219676 3',
            mRNA sequence.
ACCESSION AA687846
VERSION   AA687846
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1. (bases 1 to 421)
JOURNAL   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-f@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            DNA Sequencing by: Greg Lennon, Ph.D.
            Clone distribution: Washington University Genome Sequencing Center
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES             Location/Qualifiers
     source           1..421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1219676"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT73 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   128 a   71 c   69 g   153 t
ORIGIN
Query Match      3.4%; Score 181; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.3e-68;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1805 gcgtgaataactaagaagttgggcaaccccaacagccctacaaatcttctcgtggaat 1864
Db 241 GCGTGAATAACTAAGAAGTTGGGCAACCCCAACAGCCCTACAAATCCTTCTCTGGAAAT 300
QY 1865 gatcaattctctcttgagaggatagcacctgtgcacatagataccgaatctatcagtgc 1924
Db 301 GATCAAGTTTCTCTTGAGAGGATAGGACCTGTGCACATAGATACCGAATCTATCAGTGC 360
QY 1925 tcttattaacaagtgaacaaatcaatagatggcaacgacgagatgatgaagatgagggtgt 1984
Db 361 TCTTATTAAACAAGTGAACAAATCAATAGATGGAAACAGCAGCATGATGAAGATGAGGGTGT 420
QY 1985 t 1985
Db 421 T 421

```

```

RESULT 46
AI140791/C
LOCUS   AI140791
DEFINITION O015a05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1566224 3', mRNA sequence.
ACCESSION AI140791
VERSION   AI140791
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 306)
JOURNAL   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 1585 Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 230.

FEATURES             Location/Qualifiers
     source           1..306
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1566224"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and clones: Soares NBHSF pool 1:
                     309384-310919, 323208-325895 Soares NB2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares NB2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NBHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NBHOT
                     pool 1: 723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT   110 a   43 c   39 g   114 t
ORIGIN
Query Match      3.1%; Score 164; DB 16; Length 306;
Best Local Similarity 99.5%; Pred. No. 6.7e-61;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5003 tgataggaggacaaccatgcaaatgtgaaatagctcgaagtcttggtactttaca 5062
Db 272 TGATAGGAGGACAACCATGCAAAATGTGAAATAGTCTGGAAGTCTTGGATTACTTTACA 213
QY 5063 cctcagattgattgtcccaagaatttctgtgcttctcatggaatgcaatttaagaa 5122
Db 212 CCTCAGTATTGATTGTGCCAGAAATTTTTCGCCTTTCATGCAATGAAATTTTAAGAA 153
QY 5123 gaaagattaaagcatttttaatttaagagtgctgtgtatataaaaaataagtactgaattctt 5182
Db 152 GAAAGATTAAAGTATTTTAAATTTTAAAGAGTGTGTATAAAATAAATGACTGAAATCTCT 93
QY 5183 tatccattttatcatctcttcagtttttattat 5217
Db 92 TATCCCATTTTATCATCTCTTCAGTTTATTATAT 58

```



```

RESULT 43
AA747568      249 bp      mRNA      22-JAN-1998
LOCUS      nx92c01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269696 3',
DEFINITION      mRNA sequence.
ACCESSION      AA747568
VERSION      AA747568.1 GI:2787526
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 249)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrrp/image/image.html
Insert Length: 532 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 190.
Location/Qualifiers
1. .249
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1269696"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/Note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTT-3',
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      61 a      33 c      46 g      109 t
ORIGIN

Query Match      3.8%; Score 198; DB 11; Length 249;
Best Local Similarity 99.6%; Pred. No. 8.2e-76;
Matches 248; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4666 ttttttaagtgttcgattctgaagtcgtgtatagctttttatctcggttttaact 4725
Db 1 TTTTAAAGTTTCGATTCTCGAAGTGCCTGTATAGCTTTTATCTCGGCTTTAAACT 60

QY 4726 gacagtcaccgactgtttatggatctattgttgaaagattgttagatagatct 4785
Db 61 GACAGTACCGGACTGTTATGAGTCTATTGATTGAAAGAAATTGTTAGGATAGACT 120

QY 4786 taagcagtaattctgctgattgttattgtattttctgcaatttttactgtgaaaaaaa 4845
Db 121 TAAGCAGTAATCTGTCAGTGTGTTGTTATGTTATCTCTGCAATTTTACTGTGAAAAAAA 180

QY 4846 tttgttttcaacaattgggtgcattttttctgtgctcattttgttgagagtaaatgg 4905

```

```

Db 181 TTTGTTTCAACAATGGGTGCATTTTCTGTGATGTCACATATTGTTGGAGAGTAAATGG 240
QY 4906 tctctctccc 4914
Db 241 TCTCTTCCC 249

RESULT 44
AI572349/c      186 bp      mRNA      EST      13-APR-1999
LOCUS      te40c12.x1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:2089174
DEFINITION      3', mRNA sequence.
ACCESSION      AI572349
VERSION      AI572349.1 GI:4535723
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 706 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 184.
Location/Qualifiers
1. .186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2089174"
/clone_lib="Soares_NhMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/Note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      78 a      23 c      21 g      64 t
ORIGIN

Query Match      3.5%; Score 183; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.1e-69;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5069 tattgattgtcccgagaattttctgccccttcattggccaatgaaattttaagaagaaga 5128
Db 186 TATTGATTTGTCCCGAATTTTCTGCCCTTTCATGCAATGAAATTTTAAAGAAGAAGA 127

QY 5129 tttaaagtatttaatttaaagagtggttatataaaatgaatgtactgaattcttcc 5188
Db 126 TTTAAGTATTTTAATTTTAAGAGTGTGTATATAAATATGATCTGATCTTTATCCC 67

QY 5189 attttatcatcctttcagttttttatttaactctactgtatcaataaaattctgtaattgaa 5248
Db 66 ATTTTATCATCCTTTCAGTTTTTATTAACTACTGTATCAATAAAATTTCTGTAATTGAA 7

QY 5249 tga 5251

```











```

FEATURES
source
Location/Qualifiers
1. .446
/organism="Homo sapiens"
/db_xref="CDB:3848693"
/db_xref="taxon:9606"
/clone="IMAGE:220700"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
```

BASE COUNT 118 a 69 c 97 g 153 t 9 others

ORIGIN

```

Query Match 4.4%; Score 231; DB 158; Length 446;
Best Local Similarity 99.6%; Pred. No. 2.7e-90;
Matches 281; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaatggagctgtgaagagtgagcagttggagccttacttggtagcccccatacatt 4550
Db 1 CACAAAATGGAGCTGCTGAAGAGTGGACAGTTGGACCTTACTTTGGTGCACCCCATACATT 60

QY 4551 tgtgtcacatgctttagccatacacatggttaacattgactatgagcttctgtgaagtg 4610
Db 61 TGTGTCACATGCTTTAGCCATACACATGTTAATGACTATGAGTCTGTGAAGTG 120

QY 4611 taatgtcgatggtctatgtagacataaagaagaactgttaatatctttttttttt 4670
Db 121 TAATGTGCGATGGCTATGTAGACATAAAGAAGAACTGTGAATATCTTTTCTTTT 180

QY 4671 ttaatgttctgattctgaagtctgtatagcttttatctgcggtttaaacctgacag 4730
Db 181 TTAATGTTCTTGATTCNGAAGTCTGTATAGCTTTTATCTGCGGCTTTAAACTGACAG 240

QY 4731 taccgcagctgttattggatctattgattgaaagaatttg 4772
Db 241 TACCGACTGTTTATGGATCTATTGATTGAAAGAAATTG 282

RESULT 38
BF222521 464 bp mRNA EST 09-NOV-2000
LOCUS BF222521
DEFINITION 7p53q08.x1 NCI-CCAP-Pr28 Homo sapiens cDNA clone IMAGE:3649719 3'
similar to TR:Q9Y451 Q9Y451 ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
SHUTOFF ASSOCIATED PROTEIN. ; mRNA sequence.
ACCESSION BF222521
VERSION BF222521.1 GI:11129698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

```

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
High quality sequence stop: 412.  
Location/Qualifiers  
1. .464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3649719"  
/clone\_lib="NCI-CCAP-Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CCAP-Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneIDs  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 142 a 82 c 78 g 162 t

ORIGIN

```

Query Match 4.2%; Score 222; DB 146; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.4e-86;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1805 gcgtgaaataactaagaagttggcaccacacagcctacaaatccttctcctggaaa 1864
Db 239 GCGTGAATAACTAAGAAGTTGGCAACCCCAACAGCGCTCAAAATCCTTCTTGGAAAT 298

QY 1865 gatcaagttctcttgagagagatagcacctgtgcacatagataccgaatctatcagtcg 1924
Db 299 GATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGC 358

QY 1925 tcttattaaacaagtgaacaaatcaatagatggaacagcagatgatgaagatgaggtgt 1984
Db 359 TCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGT 418

QY 1985 tccaactgatcaagccatcagacgaggtcttgaactgcttaa 2026
Db 419 TCCAACATGATCAAGCCATCAGAGCAGGCTTTGAACTGCTTAA 460

RESULT 39
AA333118
LOCUS AA333118
DEFINITION EST37155 Embryo, 8 week I Homo sapiens EST 21-APR-1997
ACCESSION AA333118
VERSION AA333118.1 GI:1985362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.J., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Palligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

```



```

/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: parathyroid gland; Vector: pT7p3D (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer
[5'-TGTTACCACTGAAGTGGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTT
TTTTT-3'], double-stranded cDNA was size selected, ligated
to Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7p3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
BASE COUNT      95 a      36 c      33 g      93 t
ORIGIN

Query Match      4.6%; Score 240; DB 12; Length 257;
Best Local Similarity 100.0%; Pred. No. 3.1e-94;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5011 ggacaccatgcgaattgtgaatagctcgaagttcttggtattctaccacctcagta 5070
Db 257 GGACAACCACTGCAAAATGTGAAATAGTCTCTGGAATTTCTTGGATTACTTTACACCTCAGTA 198

QY 5071 ttgattgtccagaattttctggccttcctcgaatgcaattgaaatttaagagaagaagatt 5130
Db 197 TTGATTGTGCCAGAAATTTCTGGCCCTTCATGGCAATGAAATTTTAAAGAAGAAAGATT 138

QY 5131 taaagtatttttaagagtggtgtataaaataatgtactgaattctttatcccat 5190
Db 137 TAAAGTATTTTAAATTTTAAAGAGTGTGTTATAAATAATGACTGAATTCCTTTATCCCAT 78

QY 5191 ttatcatcttcagtttttattatcatctactgtatcaataaaattctgtaatttgaatg 5250
Db 77 TTTATCATCTTTCAGTTTTTATTATTAATCTACTGTATCAATAAATTTCTGTAATTGGAATG 18

RESULT 36
BE551003
LOCUS
DEFINITION      7b67b03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3233261 3',
rna sequence.
ACCESSION      BE551003
VERSION        BE551003.1 GI:9792695
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        1 (bases 1 to 535)
TITLE          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT        Tumor Gene Index
               Unpublished (1997)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
               Emmert-Buck, M.D., Ph.D.
               cDNA Library Preparation: M. Bento Soares, Ph.D.
               cDNA Library Arrayed by: Greg Lennon, Ph.D.
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL, send email to:
               info@image.llnl.gov
               Seq primer: -40UP from Gibco
               High quality sequence stop: 475.
               Location/Qualifiers

```

```

1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3233261"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/Note="Organ: lung; Vector: pT7p3D-Pac (Pharmacia) with a
modified polylinker; plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      183 a      90 c      108 g      153 t
ORIGIN

Query Match      4.5%; Score 237; DB 137; Length 535;
Best Local Similarity 100.0%; Pred. No. 6.2e-93;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2919 aaggagagagagctcattaggaattgttggtgaaataataataagcgaggag 2978
Db 167 AAGGAGAGAGAGCTCATGCTAGGCAATGTTGGTGAATAATAAATGTAAAGCGGAG 226

QY 2979 tatctgaagcagcatcagctgttagtgaataattgtctctctaccagagatgtt 3038
Db 227 TATCTGAACGAGATGACGCTGTAGTGAATAATTTATGTCCTCTTACCAGAGTAGTT 286

QY 3039 gttccatatataattccatttggcaccatgacccagatttatgtcaaaagtacaggatatt 3098
Db 287 GTTCCATATACAAATTCACCTTTTGGCACATGACCCAGATATTGCTCAAAGTACAGGATATT 346

QY 3099 gaacaactaaagatgttaagaagaattcttgggttcttggaataataataagct 3155
Db 347 GAACAACCTTAAAGATGTTAAAGAAATCTCTTGGTTTGTCTGGAATAATTAATGGCT 403

RESULT 37
H93424
LOCUS
DEFINITION      H93424 446 bp mRNA EST 01-DEC-1995
IMAGE:220700 3', mRNA sequence.
ACCESSION      H93424
VERSION        H93424.1 GI:1099752
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        1 (bases 1 to 446)
TITLE          Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
               M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Narra, M., Parsons, J.,
               Rifkin, L., Rohlfing, F., Soares, M., Tan, F., Trevisan, E., Waterston
               R., Williamson, A., Wohlmann, P. and Willson, R.
               The WashU-Merck EST Project
               Unpublished (1995)
JOURNAL        Contact: Willson RK
COMMENT        Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@wustl.edu
               High quality sequence stops: 239
               Source: IMAGE Consortium, LLNL
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Insert Length: 1379 Std Error: 0.00
               Seq primer: Promega -2ml3
               High quality sequence stop: 239.

```



1414920-1417991 and 1520904-1522439}. Subtraction by Bento Soares and M. Fatima Bonaldo.  
BASE COUNT 92 a 59 c 71 g 135 t  
ORIGIN

Query Match 5.3%; Score 278; DB 164; Length 357;  
Best Local Similarity 99.7%; Pred. No. 6.4e-111;  
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4707 ttatctgcgcttaactgacagtagccgactgtttatctggatctattgattgaaaag 4766  
Db 29 TTATCTCGCGCTTTAAACTGACAGTACCGACTGTTTATTGGATCTATTGATTGAAAAG 88  
QY 4767 aatttgtagatagattcaagcagtaactctgcaatctgcaatctgatttatttatttctgca 4826  
Db 89 AATTGTAGATAGATCTTAAAGCAGTAATCTGCACTGTTTGTATTGTTTGTATCTTGCA 148  
QY 4827 atttactgtgaaaaaaattgttttcaacaattggttcatttcttctgattgcaactat 4886  
Db 149 ATTTTACTGTGAAAAAAATTTGTTTCAACAATGGTGTCAATTTCTTGTATGTCACAT 208  
QY 4887 ttgttgagagttaaatggtctctccctttgtgtgtattctacactagtttactcctggg 4946  
Db 209 TTGTTGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACCTAGTGTTTACTCTCTGG 268  
QY 4947 cacccttaattcagagggtcaaaattgtctgccattacacagagaggtgcctctgat 5006  
Db 269 CACCTTAATCTCAGAGGTGCTAAATGTCTGCCATTACACGAGGATGCCCTCTGAT 328  
QY 5007 agggaggaacaccatgcaaaattgtgaata 5035  
Db 329 AGGAGGACACCATGCAAAATGTGAATA 357

RESULT 34  
BG387646  
LOCUS 602412496F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4521095 5',  
DEFINITION mRNA sequence.  
ACCESSION BG387646  
VERSION BG387646.1 GI:13281092  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 802)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: AFCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1A10419 row: p column: 24  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .802  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4521095"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SF0RM6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library.  
BASE COUNT 270 a 149 c 167 g 216 t  
ORIGIN

Query Match 5.1%; Score 269; DB 153; Length 802;  
Best Local Similarity 99.7%; Pred. No. 5.6e-107;  
Matches 389; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1158 gctattagacatgatgttattgtgtcaatagttacagctgtcaaaaggatattcttctg 1217  
Db 1 GCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGTAAAGAGATATTCTTCTG 60  
QY 1218 gtcaatgatcacttacttaattttgttgagagagacattagacaaacgagtgagagta 1277  
Db 61 GTCAATGATCATTACTTACTTAAATTTTGTGAGAGAGAACATTAGACAAACGATGGAGTA 120  
QY 1278 cgcaaaagaagccatgatgggacttgcacaaatttataagaaatgctttacagtcaag-c 1336  
Db 121 CGCAAAAGAACCCATGATGGGACTTGCCTCCAAATTTATAGAAATATGCTTTACAGTCAGAC 180  
QY 1337 agctggaaaagatgctgcaaaacagatagcatgatcaaaacaaattgtcacatatata 1396  
Db 181 AGCTGGAAGATGCTGCAAAACAGATAGCATGATGATCAAAACAAATTCGTACATATA 240  
QY 1397 ttatcaaaatagattatgatgactacttctgttgacggatctttgctcaatcattgt 1456  
Db 241 TTATCAAAATAGTATGATGATCGACTACTTGTGTAACGGATCTTGTCTCAATACATGGT 300  
QY 1457 tcctcacaatttagaaactacagacgagtgaaatgcttattacttctgtatgccacact 1516  
Db 301 TCCTCACAAATTTAGAACTACAGAACGGATGAATGCTTATATTACTTGTATGCCACACT 360  
QY 1517 ggatttaaatgctgtgaaagcattgaatga 1546  
Db 361 GGATTTAAATGCTGTGAAGCATTGAATGA 390  
RESULT 35  
AA854212/c  
LOCUS a172a03.s1 Soares.parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone  
DEFINITION IMAGE:1401964 3', mRNA sequence.  
ACCESSION AA854212  
VERSION AA854212.1 GI:2941750  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 257)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 554 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 182.  
Location/Qualifiers  
1. .257  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1401964"











REFERENCE 1 (bases 1 to 410)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
 JOURNAL Tumor Gene Index  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbbrp/image/image.html](http://www-bio.llnl.gov/dbbrp/image/image.html)  
 Insert Length: 719 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 398.

## FEATURES

source  
 1..410  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1603350"  
 /clone\_lib="NCI-CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
 112 a 59 c 88 g 151 t

BASE COUNT  
 ORIGIN

Query Match 5.9%; Score 311; DB 14; Length 410;  
 Best Local Similarity 99.7%; Pred. No. 2.1e-125;  
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4458 ggtcaagcttgaggctgaataaagcctttgatgcacaaaatgggactgctgaagatgga 4517  
 |||||  
 Db 1 GGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGAGTGA 60

QY 4518 cagttggaccttacttggtagcccatcatttgggtcacatgcttttagccatacaca 4577  
 |||||  
 Db 61 CAGTTGGACCTTACTTTGGTGACCCCATACGTTTGGTGCACATGCTTTAGCCATACACA 120

QY 4578 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtctatgtagacataa 4637  
 |||||  
 Db 121 TGGTAACATTGACTATGAGTCTTGTGAAAGTGTAAATGTGCGATGCTATGTAGACATAA 180

QY 4638 aqaagaacttgaataatccttttttcttttttaagtcttctgaattctctgaagtgcctt 4697  
 |||||  
 Db 181 AGAAGAACTTGTAAATATCTTTTCTTTTAAATGTTTCTGATTCTCTGAAGTGCTT 240

QY 4698 gtatagcttttctgcggctttaactgacagtcacccagctgtttattggatctattga 4757  
 |||||  
 Db 241 GTATAGCTTTATCTCGCGCTTTAACTGACAGTACCCGACTGTTTATTTGGATCTATGA 300

QY 4758 ttgtgaaagaattgttaggatatagatctttaagcagtaactctgcagtggtttgtattgta 4817  
 |||||  
 Db 301 TTTGAAAGAAATTTGTAGGATAGATCTTTAAGCAGTAATCTGTCAGTGTGTATTGTGA 360

QY 4818 tt 4819  
 ||  
 Db 361 TT 362

RESULT 29  
 AA911125

## LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

EMAIL

TISSUE

PROCUREMENT

CDNA

LIBRARY

PREPARED

BY

CLONE

DISTRIBUTION

NCI-CGAP

CLONE

LIBRARY

NCI-CGAP

LU5

TISSUE

TYPE

"carcinoid"

LAB

HOST

"DH10B"

NOTE

"Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN

Query Match 5.8%; Score 305; DB 13; Length 356;  
 Best Local Similarity 99.7%; Pred. No. 9.2e-123;  
 Matches 355; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4457 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtctatgtagacata 4536  
 |||||  
 Db 1 TGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGAGTGG 60

QY 4517 acagttggaccttacttggtagcccatcatttgggtcacatgcttttagccatacaca 4576  
 |||||  
 Db 61 ACAGTTGGACCTTACTTTGGTGACCCCATACGTTTGTGTCACATGCTTTAGCCATACAC 120

QY 4577 atcggttaacattgactatggagctcttgaaagtgtaatgtgcgactggtctatgtagacata 4636  
 |||||  
 Db 121 ATGGTAACATTGACTATGAGTCTTGTGAAAGTGTAAATGTGCGATGCTTATGTAGACATA 180

QY 4637 aagaagaacttgaataatccttttttcttttttaagtcttctgaattctctgaagtgcct 4696  
 |||||  
 Db 181 AAGAGAACTTGTAAATATCTTTTCTTTTAAATGTTTCTGATTCTCTGAAGTGCT 240

QY 4697 tgtatagcttttctgcggctttaactgacagtcacccagctgtttattggatctattg 4756  
 |||||  
 Db 241 TGTATAGCTTTATCTCGCGCTTTAACTGACAGTACCCGACTGTTTATTTGGATCTATTG 300

QY 4757 atttgaagaagaatttggtaggatatagatctttaagcagtaactctgtcagtggttattg 4812



```

/db_xref="taxon:9606"
/clone="IMAGE:4609533"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1: Sfil (ggccattggcc); Site_2: Sfil (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGCG-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT      225 a   129 c   144 g   231 t
ORIGIN

```

```

Query Match      5.9%; Score 313; DB 153; Length 729;
Best Local Similarity 99.7%; Pred. No. 2.8e-126;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 470 tgcctgggtcaagctatataacatagcttgggttagagatagcaatgaaattttcac 529
|||||
Db 42 TGCTGGGTCAAGTCAATATAACATATGCTTTGAGTTAGAAAGATAGCAATGAAATTTTCAC 101

QY 530 ccagctatacagaacaccttttcagttataaacaatggccacaatcagaagctccatat 589
|||||
Db 102 CCAGCTATACAGAACCTTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATAT 161

QY 590 gcacatggtagacacttatgagctctattattgtgaaggtgatacagtgctcaggagct 649
|||||
Db 162 GCACATGGTAGACCTTATGAGCTCTATTATTGTGAAGGTGATACAGTGTCTCAGGAGCT 221

QY 650 ttggatacaggttttagtaaatctggtacctgctcacaagaattaaacaagaagcata 709
|||||
Db 222 TTTGGTACACGGTTTATAGTAATCTGTGTACCTGTCTCAAGAATTTTAAACAAGCAGATA 281

QY 710 tgattggcaaggcttactgaagggacagctcaagctatigagccatatattaccac 769
|||||
Db 282 TGATTTGGCAAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGTGAGCCATATATTACAA 341

QY 770 ttttttaacaggttctgattggtggaaacacatctatcagcgattttgtcagagcattg 829
|||||
Db 342 TTTTAAATCAGGTTCTGATGCTTGGGAAACATCTATCAGCGATTGTGTCAGAGCATGT 401

QY 830 cttt 833
|||||
Db 402 CTTT 405

```

```

RESULT 27
AI299646      406 bp      mRNA      EST      29-JAN-1999
LOCUS      gn12a10.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898010 3',
DEFINITION      mRNA sequence.
ACCESSION      AI299646
VERSION      AI299646.1 GI:3959196
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 406)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.

```

```

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 548 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 394.

```

```

FEATURES
Location/Qualifiers
source
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1898010"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      112 a   58 c   87 g   149 t
ORIGIN

```

```

Query Match      5.9%; Score 311; DB 18; Length 406;
Best Local Similarity 99.7%; Pred. No. 2.1e-125;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4458 ggtcaagcttgagctgaataaagccttgatgcacaaaatggagactgctgaagtgga 4517
|||||
Db 1 GGTCAGCTTGGAGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGAGTGGA 60

QY 4518 cagtgtgaccttacttgggtgaccccatcacatttgggtcacatgcttttagccatacaca 4577
|||||
Db 61 CAGTTGGACCTTACTTTGGTGACCCCATACGTTTGTGTCACATGCTTTAGCCATACACA 120

QY 4578 tggtaacattgactatggagctcttggtgaagtgtaattgctgagatggctatgacacataa 4637
|||||
Db 121 TGCTAACATTGACTATGGAGTCTTGTGAAAGTGTAATGTGCGATGGCTATGTAGACATAA 180

QY 4638 agaagaacctgttaaatatcttttcttttttttttttttttttttttttttttttttttt 4697
|||||
Db 181 AGAAGAAACCTGTAAATATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 240

QY 4698 gtatagcttttatctgcggcttttaaaactgacagtcaccgactgtttattggtattga 4757
|||||
Db 241 GTATAGCTTTTATCTGCGGCTTTAACTGACAGTACCGACTGTTTATGATCTATGTA 300

QY 4758 ttgaaagaatttggtagatagatcttaagcagtgtaattctgctagctgtttgtattgta 4817
|||||
Db 301 TTTGAAAGAAGATTGTTAGGATAGATCTTAAGCAGTAATCTGTGCTGCTTTGTTATTGTA 360

QY 4818 tt 4819
|||||
Db 361 TT 362

```

```

RESULT 28
AA987361
LOCUS      or92904.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1603350 3',
DEFINITION      mRNA sequence.
ACCESSION      AA987361
VERSION      AA987361.1 GI:3172725
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```



```

/clone_lib="Stratagene HeLa cell s3 937216"
/dev="female"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      112 a      81 c      64 g      133 t      1 others
ORIGIN

Query Match      6.7%; Score 354; DB 10; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-144;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2888 aatctgtgcccttctgcaaaagatcctgaaaggagagagagatcgttaggcaatg 2947
      |||||||
Db 354 AATCTGTGCCCTTTGTGCAAAAGATCCTGTAAGGAGAGAGAGAGCTCATGCTAGGCAATG 295

QY 2948 ttgtgtaaaataataatgaaggcggagtagtctgaagcagcatgcagctgttagtga 3007
      |||||||
Db 294 TTTGGTGAATAATATAAATGTAAGCGGAGATATCTGAAGCAGCATGCAATTCACCTTTTGGCACA 235

QY 3008 aaattattgtctctaccagagatgtgttccatatacaattcaccttttggcaca 3067
      |||||||
Db 234 AAAATATTGTCTCTCTACAGAGATGTGTTCATATACAAATTCACCTTTTGGCACA 175

QY 3068 tgaccagattatgtaaaagtcagagattgaacaacttaagatgtttaagaatgtct 3127
      |||||||
Db 174 TGACCAGATTATGTCAAAAGTACAGATATTGAACAACTTAAAGATGTTTAAAGATGTCT 115

QY 3128 ttgggttctcgaaattattagctgaaataaataaataaagtcagcttttatcag 3187
      |||||||
Db 114 TTGGTTGTCTGGAATAATTAATGGCTAAAATGAATGAATAACAGTCACGCTTTTATCAG 55

QY 3188 aaagatggttagaaaaattataacaaacaaagatgcccagagcagatgatgc 3241
      |||||||
Db 54 AAAGATGTTAGAAAATATTAAACAAACAAAGATGCCCAAGGACAGATGATGC 1

RESULT 25
LOCUS      AV710229      738 bp      mRNA      EST      09-OCT-2000
DEFINITION AV710229 Cu Homo sapiens cDNA clone CuAAJC12 5', mRNA sequence.
ACCESSION  AV710229
VERSION     AV710229.1 GI:10728785.
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 738)
AUTHORS    Peng,Y., Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J.,
Chen,Z. and Han,Z.
TITLE      Homo sapiens cDNA Cu clones
JOURNAL     Unpublished (2000)
CONTACT    Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzgchgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES    Location/Qualifiers
            source      1..738
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"

```

```

/clone="CuAAJC12"
/clone_lib="Cu"
/tissue_type="adrenal cortico adenoma for Cushing's syndrome"
/dev_stage="Adult"
/notes="Vector: pBluescript sk(-)"
BASE COUNT      221 a      113 c      150 g      253 t      1 others
ORIGIN

Query Match      6.1%; Score 324; DB 32; Length 738;
Best Local Similarity 99.7%; Pred. No. 4.2e-131;
Matches 374; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4      gagagcccgagtagcgagtagcgagtcggaacccggagggtagaataattctg 63
      |||||||
Db 2      GAGAGCCCGGAGTAGCGGAGTAGCGAGTCGCGCAACCCGGAGGGTAGAAAATATTTCTG 61

QY 64      tcattgctcatcaagactagaccatgatggaaaaattacatatccgcctggggtca 123
      |||||||
Db 62      TCATGCTCATTCAAAGACTAGGACCAATGATGGAATAATACATATCCGCTGGGTCA 121

QY 124      aggaataatcagataaaatatctaaagagagagatggtgagacattaaagatggttga 193
      |||||||
Db 122      AGGAATAATCAGATAAAATATCTAAAGAGAGAGATGTGTGAGACATTAAAGATGTTGTGA 181

QY 184      aaactttttagatgagcagagactctgaagaagaagaagagagctttatttaaacctag 243
      |||||||
Db 182      AAATTTTATGATATGACAGGACTCTGAAGAAGAAAGAGAGCTTTATTTAAACCTAG 241

QY 244      cttacattgttctcagatatttttctcaagcattcctgtaagatgttcgcttactgg 303
      |||||||
Db 242      CTTTACATCTTGTCTCAGATTTTTTCTCAAGCATCTCTGATATAAGATGTTCGCTTACTGG 301

QY 304      taacctgtccttctgctgatttttcagatttatctcctaagctccttaacatccc 363
      |||||||
Db 302      TAGCTGCTGCGCTTGTCTGATATTTTTCAGGATTTATGCTCTCTGAAGCTCCTTACATCCC 361

QY 364      ctgataaaactaaagg 378
      |||||||
Db 362      CTGATAAACTAAAGG 376

RESULT 26
LOCUS      BG432960      729 bp      mRNA      EST      14-MAR-2001
DEFINITION BG432960 NTH_MGC_75 Homo sapiens cDNA clone IMAGE:4609533 5', mRNA sequence.
ACCESSION  BG432960
VERSION     BG432960.1 GI:13339466
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 729)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LLCM1353 row: e column: 22
            High quality sequence stop: 722.
FEATURES    Location/Qualifiers
            source      1..729
                        /organism="Homo sapiens"

```







```

DEFINITION 7n43b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3567291 3',
mRNA sequence.
ACCESSION BF111072
VERSION BF111072.1 GI:10940762
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 470.
FEATURES
source
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3567291"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 133 a 81 c 109 g 201 t 1 others
ORIGIN
Query Match 7.2%; Score 381; DB 144; Length 525;
Best Local Similarity 99.6%; Pred. No. 4.1e-156;
Matches 481; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4453 ttttgggtcaagcttgaggctgaataaagccttttgatgcacaaatgggactgctgaaga 4512
Db 1 TTTTGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAATGGGACTGCTGAAG 60

QY 4513 gtgacagtgacaccttaattggtgacccatacatattggtgacatgcttagccat 4572
Db 61 GTGGACAGTGGACCTTACTTTGGTGACCCCATACGCTTGTGGTCACATGCTTTAGCCAT 120

QY 4573 acacatggtlaacatgaactagatgagctctgtgaaagttaagtgcgactataga 4632
Db 121 ACATGTTGAACATGACATGAGGCTTGTGAAAGTGAATGTCGATGGCTATGTAGA 180

QY 4633 cataaagaagaacttgtaaatctctttcttttttttaattgttctgattctgaag 4692
Db 181 CATAAAGAAGAACTGTAAATATCTTTTCTTTTAAATGTTCTGATTCGTGAAG 240

QY 4693 tgcctgtatagcttttatctgcggctttaactgacagtcaccgactgtttatgatatc 4752
Db 241 TGCTGTATAGCTTTTATCTGCGGCTTAACTGACAGTACCCGACTGTTTATTGGATCT 300

QY 4753 attgatttgaagaatttcttaggatacatcttaagcagtaactctgcagtttgtat 4812
|||||

```

```

Db 301 ATTGATTGAAAAGAAATTGTTAGGATAGATCTTAAGCAGTAATCTGTCAGTCTTTGTAT 360
QY 4813 ttgtatttctgcaatttactgtgaaaaaatttgttttcaacaattgtgtcatttt 4872
|||||
Db 361 TTGTATTCTCTGCAATTACTGTGAAAAAATTTGTTTCAACAATTTGTTGTCATT 420
QY 4873 cttgactgctactatttctgtggagagtaaatggtctctcttccctttgtgtatcttaacctag 4932
|||||
Db 421 CTTGATGTCTACTATTATTGTTGGAGAGTAAATGCTCTCTCCCTTTGTGTATCTTACCTAG 480
QY 4933 tgt 4935
|||||
Db 481 TGT 483

RESULT 22
BF446410
LOCUS BF446410 390 bp mRNA EST 01-DEC-2000
DEFINITION 7p38h09.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3648232 3',
mRNA sequence.
ACCESSION BF446410
VERSION BF446410.1 GI:11511548
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES
source
1..390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3648232"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 105 a 56 c 83 g 146 t
ORIGIN
Query Match 7.0%; Score 368; DB 148; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.1e-150;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4452 tttttgggtcaagcttgaggctgaataaagccttttgatgcacaaatgggactgctgaag 4511
|||||
Db 1 TTTTGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAATGGGACTGCTGAAG 60

```







```

Db 229 AAATAGTCTGAGGTTCTTGGATTACTTTACACCTCAGTATTCGATTTGCCAGAAATTTT 170
QY 5091 ctggccttcctggaatgaaatcttaagaagaagatttaagatttttaatttttaa 5150
Db 169 CTGGCCTTTTCATGGCAATGAAATTTTAAAGAGAAAGATTTAAAGATTTTAAATTTTAA 110
QY 5151 gagggtgttataaaataagtgactgaatctctttatccatctttatcatcctttcagtttt 5210
Db 109 GAGTGTGTTATATAAATAATGTACTGAATCTCTTTATCCCATTTTATCATCCTTTTCAGTTTT 50
QY 5211 tattaatcactgtatcaataaaatctgtaaatttgaatgag 5252
Db 49 TATTAATCTACTGTATCAATAAATTTCTGTAAATTTGAATGAG 8

RESULT 18
AI681422
LOCUS AI681422 486 bp mRNA EST 16-DEC-1999
DEFINITION tx46e01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272632 3',
mRNA sequence.
ACCESSION AI681422
VERSION AI681422.1 GI:4891593
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 714 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 460.
FEATURES
Location/Qualifiers
1..486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2272632"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522435). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 125 a 73 c 101 g 187 t
ORIGIN
Query Match 7.3%; Score 384; DB 23; Length 486;
Best Local Similarity 99.6%; Pred. NO. 2e-157;
Matches 484; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4457 tggctcagcttgagctgaataagcctttgacacaaatgggactgctgaagtggtg 4516
Db 1 TGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAATGGGACTGCTGAAGAGTGG 60

```

```

QY 4517 acagttggaccttaocttttggtgaccccatatacatatttggtcacatgctcttagccatacac 4576
Db 61 ACAGTTGGACCTTACTTTGGTGGACCCCATACATCGTTTGGTGCACATGCTTTAGCCATACAC 120
QY 4577 atggttaacattgactgagagctcttgtaaaagtgaatgtgcgatggctatgtagacata 4636
Db 121 ATGGTAAACATTGACTATGGAGTCTTGTGAAAGTGTAAATGTGCGATGGCTATGTAGACATA 180
QY 4637 aagaagaactgtgaaatctcttttctttttaaagtttctctgattctcgaagtgc 4696
Db 181 AAGAAGAACTGTAAATATCTTTTCTTTTAAAGTTTCTGATTTCTGAAGTGCT 240
QY 4697 tgtatagcttttatctgcggtcttaaacctgacagtaaccogactggtttattggatctattg 4756
Db 241 TGTATAGCTTTATCTCGCGCTTTAAACTGCACAGTACCCGACTGTTTATTGGATCTATTG 300
QY 4757 atttgaagaagaatttttaggatagatcttaagcagtaactctcagtggtttattgt 4816
Db 301 ATTGTGAAAGAAATTTCTTAGGATAGATCTTAAGCAGTAATCTGTCACTGTTTGTATTGT 360
QY 4817 attttctgcaattttactgtgaaaaaaatttcttccaacaatttggtgctatttcttg 4876
Db 361 ATTCTCTGCAATTTTACTGTGAAAAAAATTTCTTTTCAACAATTTGGTGTCAATTTCTTG 420
QY 4877 atgtcactatttggtggagagtgtaaatggtctctctcccttggtgctattcaccctagttt 4936
Db 421 ATGTCACTATTTGTGGAGAGTTAAATGGTCTCTCTCCCTTTGTGTATCTTACCTAGTGTT 480
QY 4937 tactcc 4942
Db 481 TACTCC 486

RESULT 19
BF511856
LOCUS BF511856 490 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-B14-aps-d-12-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088558 3', mRNA sequence.
ACCESSION BF511856
VERSION BF511856.1 GI:11595154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
Location/Qualifiers
1..490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3088558"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; NCI_CGAP_Sub8
is a subtracted library derived from NCI_CGAP_Sub5. The
NCI_CGAP_Sub8 library had 2.5 million recombinants. A

```



IMAGE:1627886 3', mRNA sequence.  
 ACCESSION A1016896  
 VERSION A1016896.1 GI:3231232  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 461)  
 AUTHORS NCI-CCAG <http://www.ncbi.nlm.nih.gov/ncicag>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
 Insert Length: 548 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 420.  
 FEATURES  
 Location/Qualifiers  
 1..461  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1627886"  
 /lab\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pTT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI CGAP GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 182 a 76 c 72 g 131 t  
 ORIGIN  
 Query Match 7.7%; Score 406; DB 15; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-167; Indels 0; Gaps 0;  
 Matches 406; Conservative 0; Mismatches 0;  
 Qy 4821 tctgcaatttactgtgaaaaaatttcttcaacaattggtgtcatttcttgatg 4880  
 Db 436 TCTGCAATTTTACTGTGAAAAAATTTGTTTCAACAATGGTGTCATTTCTTGATGT 377  
 Qy 4881 cactattgttggagagtaaatggctctcccttcttgatcttactagtggtttact 4940  
 Db 376 CACTATTTTGGAGAGTAAATGGTGCTCTCCCTTTGTGTATCTTACTAGTGTACT 317  
 Qy 4941 cctggccaccttaattctcagagtgctaaattgtctgcatcacacagaagatgcc 5000  
 Db 316 CCTGGGCACCTTAAATCTTCAGAGTGCTAAATGTCTGCCATTACACCAAGAGATGCC 257  
 Qy 5001 tctgtaggaggaacaaccatgcaaatgtgaaatagtcctgaagttcttgattacttta 5060  
 Db 256 TCTGATAGGAGGACACCAATGTGAATAGTCTCAAGTCTCTGGATTACTTTA 197  
 Qy 5061 cactcagattgattgtccagaatttctggccttctcagcaatgaaattttaaag 5120  
 Db 196 CACCTCAGTATGTATTTCTCCAGAAATTTCTGGCCTTTTCATGGCAATGAAATTTTAA 137  
 Qy 5121 aagaaagattaaagtattttaaagtgtgtataaaaaataatgactgaattc 5180  
 Db 136 AAGAAGATTTAAAGTATTTTAAAGAGTGTGTATATAAATAATGATGACGAATTC 77  
 Qy 5181 ttattccattttatcatccttccagtttttatttaataactactgtat 5226  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 76 TTTATCCCAATTTTATCATCCCTTCAGTTTATTATTAATCTACTGTAT 31  
 RESULT 17  
 AA625961/c 409 bp mRNA EST 15-OCT-1997  
 LOCUS zu92e08.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:745478  
 DEFINITION 3', mRNA sequence.  
 ACCESSION AA625961  
 VERSION AA625961.1 GI:2538348  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 409)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin  
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 378.  
 FEATURES  
 Location/Qualifiers  
 1..409  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5933433"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:745478"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pTT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 , Inc., and primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 155 a 67 c 66 g 121 t  
 ORIGIN  
 Query Match 7.6%; Score 402; DB 9; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-165; Indels 0; Gaps 0;  
 Matches 402; Conservative 0; Mismatches 0;  
 Qy 4851 tttaacaattggtgtcatttcttcttgatgtcactatttggagagtgtaaatgtctct 4910  
 Db 409 TTTCACAATTTGGTGTCATTTCTTGATGTCTACATTTTGGAGAGTTAAATGGTCTCT 350  
 Qy 4911 tcccttggatcttactagtggtttactcctggcacccttaattctcagagtgcta 4970  
 Db 349 TCCCTTTGTGTATCTTACTAGTGTCTTCTCTGGGACCCCTTATCTTCAGAGTGCTA 290  
 Qy 4971 aatgtctgcattacacacaagatgcctctatagagacaacatcaaatgtg 5030  
 Db 289 AATTGCTGCGCATTTACACCAAGGATGCCCTCTGTAGAGACCAACCATGCAAAATG 230  
 Qy 5031 aaatagtcctgaagtcttcttgattactttcacctcagttattgttgcacgaatttt 5090  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||



## FEATURES

Location/Qualifiers

1. 578  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:300887"  
 /clone\_lib="NCI\_CGAP\_Lu24"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 203 a 97 c 112 g 163 t 3 others  
 ORIGIN

Query Match 8.0%; Score 422; DB 120; Length 578;  
 Best Local Similarity 99.8%; Pred. No. 4.2e-174;  
 Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2755 aaatcatcattagacaatatcagctatgtcattagctatcaacgatgaatgctatc 2814

|||||  
 Db 4 AATATCATCATTTAGAACAAATATACGATGTGTCATTAGTATCAACGATGAATGCTATC 63

|||||  
 QY 2815 aagtaagacaagtgttgcacgaacttcacaaagcccttccgctttacggctccac 2874

|||||  
 Db 64 AGTAGAGCAAGTGTGGCCGAGAACTTCACAAAGCCCTTCCCGTTACGGCTCCAC 123

|||||  
 QY 2875 ttgagtatatggcaatctgtgccccttctgcaaaagatctgtaaggagagaaagatc 2934

|||||  
 Db 124 TTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCTCTGTAAAGGAGAGAGGCTC 183

|||||  
 QY 2935 atgctaggcaatgttgggtgaaataataatgtaaggcggagagatctggaagcagatg 2994

|||||  
 Db 184 ATGCTANGCAATGTTTGGTGAATAATATAATGTAAAGCGGGAGTATCTGAAGCAGCATG 243

|||||  
 QY 2995 cagctttagtgaaaaattattgtctctctaccagatgattgtgtccatatacaattc 3054

|||||  
 Db 244 CAGCTTTAGTGAAAAATTTATGCTCTCTACACAGTATGTTGTCCATATACAAATTC 303

|||||  
 QY 3055 acctttggcacatgaccagattatgtcaaaagtagcagatatttgacaacttaagatg 3114

|||||  
 Db 304 ACCTTTGGCACATGACCCAGATTTATGTCAAAGTACAGGATATTGAACAACCTTAAAGATG 363

|||||  
 QY 3115 ttaaagaattgttgggttcttggaataattaatggctaaaaatgaaataaacagtc 3174

|||||  
 Db 364 TTAAGAATGTCTTTGGTTGTCTGGAATATTAAATGGCTAAAAATGAAAAATAACAGTC 423

|||||  
 QY 3175 acgctttatcagaagatggttagaaaattataacaaacaaagaatgccccaa 3227

|||||  
 Db 424 ACGCTTTATCAGAAAGATGGTAGAAAAATTTAAACAAACAAAGATGCCCAA 476

RESULT 15  
 AA563884/c

LOCUS nk19e08.s1 NCI\_CGAP\_Coll Homo sapiens cDNA clone IMAGE:101390 3',  
 DEFINITION mRNA sequence.

ACCESSION AA563884

VERSION AA563884.1 GI:2335523

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 469)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL  
COMMENT

Tumor Gene Index  
 Unpublished (1997)

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov

Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Elias Campo,

M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 868 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 365.

## FEATURES

Location/Qualifiers

1. 469

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:101390"

/clone\_lib="NCI\_CGAP\_Coll"

/tissue\_type="tumor"

/lab\_host="SOLR (kanamycin resistant)"

/note="organ: colon; Vector: Bluescript SK-; Site\_1: EcoRI

; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo

dt. Multiple colon tumors. 5' adaptor sequence: 5'

GAATTCGGCAGAG 3' 3' adaptor sequence: 5'

CTCAGATTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

BASE COUNT 184 a 81 c 71 g 133 t

ORIGIN

Query Match 7.9%; Score 418; DB 9; Length 469;

Best Local Similarity 99.8%; Pred. No. 2.4e-172;

Matches 468; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4748 gatctattgattgaaagaatttggtagtagatcttaagcagtaactgtcagtggt 4807

|||||  
 Db 469 GATCTATTGATTGAAAGAATTTGTTAGGATAGATCTTAAGCAGTAATCTGTCAGTGT 410

|||||  
 QY 4808 tgtattgtatttctgcaatttactgtgaaaaaaattgttttcaacaattgtgtc 4867

|||||  
 Db 409 TGTATTGTTCTCTGCAATTTTACTGTGAAAAAAATTTGTTTCAACAATTTGGTGTGTC 350

|||||  
 QY 4868 attttctgtatgcactatttgggagaggttaaatggctctctcccttctgtatctta 4927

|||||  
 Db 349 ATTTCTTCATGTCATATTGTTGGAGAGTTAAATGGTCTCTCTCCCTTGTGTATCTTA 290

|||||  
 QY 4928 cctagtgttactcctgggcacaccttaattcagagggtgctaaattgtctgccattaca 4987

|||||  
 Db 289 CCTAGTGTCTTACTCTGGCACCCCTTAATCTTCAGAGGTGCTAAATTTGTCGCCATTACA 230

|||||  
 QY 4988 ccagaaggatgctctgataggagacaaccatgcaaaattgtgaaatagtcctcgaagtc 5047

|||||  
 Db 229 CCAGAGGATGCTCTGTGATAGGAGACACCATGCAAAATGTGAAATAGTCTCTGAAGTTC 170

|||||  
 QY 5048 ttggattactttcacacctcagttattgttcccaagaattttctggccttctcagggcaa 5107

|||||  
 Db 169 TTGGATTACTTTACACCTCAGTATTGATTGTGCCAGAGATTTTCTGGCCTTTCATGGCAA 110

|||||  
 QY 5108 tgaataatttaagaagaagaatttaagatttttaatttaagaagatggtgtataaaaaa 5167

|||||  
 Db 109 TGAATAATTTAAGAGAGAGATTTAAAGTATTTTAAATTTTAAAGAGTGTGTATAAAAAA 50

|||||  
 QY 5168 atgactgaattctttatccattatcatcttccagttttttattaa 5216

|||||  
 Db 49 ATGTACTGAATCTTTATCCCATTTATCATCTCTTCAGTCTTTTATTAA 1

RESULT 16

AA563884/c

LOCUS AI016896

DEFINITION Ou31c08.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

17-MAR-1999

EST

MRNA

461 bp



```

11541-018"
BASE COUNT      50 a    100 c    90 g    183 t
ORIGIN

Query Match      8.0%; Score 423; DB 23; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.5e-174;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3803 tcagcgaagtcggaagaagcgccatcacggtcttcagaatctgatgaacagcagtggcctga 3862
      |||||
Db 423 TCACGGAAGTCGGAAGAAGCGCCATACGGCTTCAGATCTCATGAACAGCAGTGGCCCTGA 364
      |||||

Qy 3863 ggaagaagggctcaagaagaatatattagaanaatgaagatgaacagatagttcgcgcaaa 3922
      |||||
Db 363 GGAAACAGAGGCTCAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCGCGCAAA 304
      |||||

Qy 3923 aaagggttaagagggccgacccaccaaaccctcttggtggagggtacaccaaagaagagcc 3982
      |||||
Db 303 AAAGGGTAAAGAGGCGCGCACCCAAAACCTCTTTGGTGGAGGTACACCAAAGAAGAGCC 244
      |||||

Qy 3983 acaatgaaacctcttaaaaaaggaagcaaaaaaatctggacctccagccaccagagga 4042
      |||||
Db 243 AACAAATGAAAACCTCTTAAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGAGGA 184
      |||||

Qy 4043 ggagggaagaagaagacaaagtggaaatacggaaacagaagtccaaaagcaaacagca 4102
      |||||
Db 183 GGAGGAAGAAGAAGAAAGCAAAAGTGGAAATACGGAAACAGAAGTCCAAAAAGCAAAACAGCA 124
      |||||

```

Db	123	CCGAGTGTCAAGGAGACACAGACGAGCAGAAATCTCCGAAATCTAGTGCATTTGAATC	64
Qy	4163	cacacagtccacacacagaaaagcaggaagaccatcaaaacgcacatcaccatcacaca	4222
Db	63	CACACAGTCCACACCCACAGAAAGGACGAGGAGACCATCAAAAACGCCCATCACCNTCACA	4
Qy	4223	acc 4225	
Db	3	ACC 1	
RESULT	14		
LOCUS	AW770571		
DEFINITION	h186f11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3008877 3' similar to TR:094237 O94237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA sequence.	578 bp mRNA	04-MAY-2000
ACCESSION	AW770571		
VERSION	EST.		
KEYWORDS	human.		
SOURCE	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 578)		
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D.		

Email: cgapops-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
Possible reversed clone: similarity on wrong strand  
Seq primer: -400P from Gibco  
High quality sequence stop: 465.



```

QY 2729 gctggcacaagaaccctgtttaccatgaatcatcacattagaaacaatatatcagctatgtgc 2788
|||||
Db 417 GCTGGCACAAGAACCCTGTGTACCATGAATCATCACATTAGAACAAATATCAGCTATGTGC 476
|||||
QY 2789 attagctatcaacgatgaatgctcatcaagttaagacaagtgttggcccaagaaccttcacaa 2848
|||||
Db 477 ATTAGTCTATCAACGATGAATGCTATCAAGTAAACAAGTGTGTGGCCAGAAACTTCACAA 536
|||||
QY 2849 aggccttcccgctttacgg 2867
|||||
Db 537 AGGCCTTCCCGTTTACGG 555
|||||

RESULT 11
AW976150 738 bp mRNA EST 02-JUN-2000
LOCUS EST388259 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION AW976150
VERSION AW976150.1 GI:8167375
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 738)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnqu@tigr.org
Seq primer: Forward.
Location/Qualifiers
1..738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGN"
/note="Vector: pBluescriptSKm"
BASE COUNT 214 a 126 c 129 g 269 t
ORIGIN
|||||

Query Match 8.2%; Score 431; DB 122; Length 738;
Best Local Similarity 99.6%; Pred. No. 4.7e-178;
Matches 531; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4674 atgtttctgatttcgaagtcgttgcattatgtatgtttatctgcggcttaaacacagctac 4733
|||||
Db 1 ANGTTTCTGATTTCGAAGTCGTTGTATAGCTTTATCTCGGCTTTAACTGACAGTAC 60
|||||
QY 4734 ccgactgtttattggatctattgattgaaagaattgttagtagatagatccttaagcagt 4793
|||||
Db 61 CCGACTGTTTATTGGATCATTTGATTGTTGAAGAATTGTAGTAGATATCTTAACAGT 120
|||||
QY 4794 aatctgcagtggtttgattgtattttctgcaatttactgtgaaaaaaattgtttt 4853
|||||
Db 121 AATCTGTCAGTGTGTTGTTATTTGTTATTTCTGCAATTTTACTGTGAAAAAAATTTGTTT 180
|||||
QY 4854 caacaattggtgctatttcttgatgctactatttcttgagaggttaaatggtctcttccc 4913
|||||
Db 181 CAACAATTGGTGTCATTTCTTGTGATGTCACATTTGTTGGAGAGTTAAATGGTCTCTCC 240
|||||
QY 4914 ctttgtgatacttaccatagttttactctcctgggacccttaattcttcagaggtgctaaat 4973
|||||
Db 241 CTTTGTGTATCTTACCTAGTGTGTTACTCTCTGGCACCCTTAATCTTCAGAGGTGCTAAAT 300
|||||

```

```

QY 4974 tgctgccattacaccagaaggatgctctgtatagaggacaaccatgcgaattgtgaaa 5033
|||||
Db 301 TGCTGCCATTACACCACAGGATGCTCTGTATAGGAGGACAACCATGCATAATTTGTGAAA 360
|||||
QY 5034 tagtctcgaagtcttgattacttacacctcagtatgtattgtcccaagaattttctg 5093
|||||
Db 361 TAGTCTCGAAGTCTTGGGATTACTTTACACCTCAGTATTGATTGGCCCAAAATTTTCTG 420
|||||
QY 5094 gcctttcatggcaatgaaaattttaagaagaagatttaaaagtattttaatttaagaag 5153
|||||
Db 421 GCCTTTTCATGGCAATGAAATTTTAAGAGAGAAGATTAAAGTATTATTTTAAAGAG 480
|||||
QY 5154 tegtgtataaaataatgtactgaattctttatccctttttatccatcttccag 5206
|||||
Db 481 TGTGTTATAAAATAATGTACTGAATCTTTATCCCATTTTATCATCTCTTTCAG 533
|||||

RESULT 12
AA282190 470 bp mRNA EST 13-AUG-1997
LOCUS zs89d08.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:704631 5',
DEFINITION mRNA sequence.
ACCESSION AA282190
VERSION AA282190.1 GI:1925314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 429.
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="GDB:585435"
/db_xref="taxon:9606"
/clone_lib="IMAGE:704631"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTCTTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 169 a 101 c 95 g 105 t
ORIGIN
|||||

Query Match 8.1%; Score 426; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 7.4e-176;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3301 agagtactacatcagtttgggaatctcttaagagaccgggtactaccagctcgtttcttca 3360
|||||

```



KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@rs-research.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 464.  
FEATURES  
source  
1. 542  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2813788"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19, testis NHT, and B-cell  
NCL-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
728408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 208 a 93 c 84 g 157 t  
ORIGIN  
Query Match 8.6%; Score 453; DB 114; Length 542;  
Best Local Similarity 99.8%; Pred. No. 1.1e-187;  
Matches 503; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4715 ggctttaactgacgacacccgactgtttatggtatgtattgattgaaagaatttgg 4774  
Db 542 GGCTTTAACTGACAGTACCGGACTGTTTATGATGATCTATGATTTGAAAGAATTGGTT 483  
QY 4775 aggatagatcttaagcagtaaatctgtcagtgattgtgtattgtatttctgcaattttact 4834  
Db 482 AGGATAGATCTTAAGCAGTAATCTGTGAGTGTGTTGTTGTTATCTCTGCAATTTTACT 423  
QY 4835 gtgaaaaaaattgttttcaacaattggtgctcatttctgtatgctcaatttggga 4894  
Db 422 GTGAAAAAAATTTGTTTCAACAATTTGGTGTCTATTTCTTGTGATGTCACATTTTGTGGA 363  
QY 4895 gagttaaagtctctcccttctgtatcttaactagtgatttactcctggcaccctta 4954  
Db 362 GAGTTAAATGGTCTCTCCCTTGTGATCTTACCTAGTGTGTTACTCTGGGACCCCTTA 303  
QY 4955 atcttcagagtgctaaattgtctgccattacaccagaagatgctcctgtagaggagac 5014  
Db 302 ATCTTCAGAGGTGCTAAATGTCTGCCATTACACCAGAGGATGCCCTCTGATAGGAGAC 243  
QY 5015 aaccatgcaattgtgaatagctcctgaagtcttgattacttctacacctcagattga 5074  
Db 242 AACCATGCAATTTGTGAATAGTCTCCTGAAGTCTCTGGATTCTTACACCTCAGATTGA 183  
QY 5075 ttgtcccaaatcttctggccttctgcaatgaaatttttaagaagaattttaa 5134  
Db 182 TTGTCCCAAGATTTCTGGCCTTTCAGGCAATGAAATTTTAAAGAAGAATTTAA 123  
QY 5135 gtattttaatttaagagtggtgtataaaataatgactgaattcttattccatttta 5194  
|||||

Db 122 GTATTTTAAATTTAAAGAGTGTGTATATAAATAATGACTGACTGAATCTTTATCCCATTTTA 63  
QY 5195 tcatcctttcagtttttatttaac 5218  
|||||  
Db 62 TCATCCCTTTTCAGTTTATTATTAAC 39  
RESULT 10  
BE873840 848 bp mRNA EST 20-OCT-2000  
LOCUS 601483992F1 NIH\_MGC\_69 Homo sapiens cDNA clone IMAGE:3886511 5',  
DEFINITION mRNA sequence.  
ACCESSION BE873840  
VERSION BE873840.1 GI:10322616  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
TITLE NIH-MGC http://mgc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@rs-research.nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLN9663 row: g column: 24  
High quality sequence stop: 654.  
FEATURES  
source  
1. 848  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3886511"  
/clone\_lib="NIH\_MGC\_69"  
/tissue\_type="large cell carcinoma, undifferentiated"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."  
BASE COUNT 261 a 163 c 191 g 232 t 1 others  
ORIGIN  
Query Match 8.3%; Score 439; DB 141; Length 848;  
Best Local Similarity 100.0%; Pred. No. 1.4e-181;  
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2429 ggtagctacttcttctgaaagattcttcatgaatgctgctccaggagaaagac 2488  
Db 117 GGTAGCTACTTCTTCTTCTGAAAGATCTTCTCATGAATGATCGGCTTCAGGGAAGAGAC 176  
QY 2489 aactaaacttgggtccagatgaagaagtatctcctgagacaaatggtcaaaattcaggc 2548  
Db 177 AACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGTCAAAATTCAGC 236  
QY 2549 tattaaatgatgggttcgactggtcacttggaatgaaataataatcacagtaatacaggac 2608  
Db 237 TATTAAATGATGGTTCGATGGCTTACTTGGAAATGAAATAATCACAGTAATCAGGAAC 296  
QY 2609 ttctaccttaagtgtctaaacaataattgcatgtagtgagacttgacagacacagg 2668  
Db 297 TTCTACCTTAAAGATTGTTTACACAACAATATTGCATGATGAGACTTGACAGACAGGG 356  
QY 2669 gaaaattagtaaacacagatgtcacgctctgagacttgcctgctggagtgctgagtgctattgtgaa 2728  
Db 357 GAAAAATTAGTAAACACAGATATGTCAGCTCTGAGACTTGTCTGCTGGAGTGTCTATTGTGAA 416  
|||||



NCI CCRP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Ronaldo.

**BASE COUNT  
ORIGIN**

Query Match	8.7%	Score 460;	DB 23;	Length 480;
Best Local Similarity	100.0%;	Pred. No. 9e-191;		
Matches 460;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 3525	actgtaagcaatgcaagcagcagctcaaatccaagctctccttggaagaataaaagggaggg	3584		
Db 480	ACTGTAAGCAATCAACGACGAGCTCAAAATCCAAGCTCTCTCGAAGAATAAAGGGGAGG	421		
QY 3585	cttgatagtctgaaatggatcacagtgaaatgaaatgaaattacacaatgtcttcacaccttg	3644		
Db 420	CTTGATAGTTCTGAAATGGATTCACAGTGAATAAGCAATTAACAAATGTCTTCACCTTTG	361		
QY 3645	ccgggggaaaaaagtgcacagagagacgactctgatcttgttaaggtctggaattggagaag	3704		
Db 360	CGGGGGAAAAAAGTGACAAGAGACGACCTCTGATCTTGTGAAGTCTGAATTTGGAGAG	301		
QY 3705	cctagaggcagggaaaaaacgcccgtcacagaacagaggagagaataatggtatggtatgac	3764		
Db 300	CCTAGAGGCGAGGAAAAAAGCCCGTCCACAGACAGAGGAGGAGAAATTTAGGTATGGATGAC	241		
QY 3765	ttgactaaagtgtgtacaggaacagaaacctaaagcgagtcagcgaaatcgcgaaagagggc	3824		
Db 240	TTGACTTAAGTTGTTACAGGAAACAGAAACCTTAAGGCGAGTCAGCGAAGTCGGAAAAAGGCG	181		
QY 3825	catacggcttcagaatctgatgaacagcagtgcccttgaggaaagaggctcaaaagaagat	3884		
Db 180	CATACGGCTTTCAGAAATCTGATGAACACGACGTGGCGCTCAGGAAAAAGAGGCTCAAAAGAGAT	121		
QY 3885	atattagaaaatgaagatgaacagaaatgctccgccaataaagggttaaaagagggccgacc	3944		
Db 120	ATATTAGAAANTGNAGATGAACAGAAATAGTCCGCCCAAAAAGGGTTAAAGAGGCCGACCA	61		
QY 3945	ccaaaacctctggtggaggtacaccaaaagaagagccaa	3984		
Db 60	CCAAAACCTCTTGTTGGAGGTACACCAAAAAGAGAGCAAA	21		

RESULT	8	
BF002153/c		
LOCUS	540 bp	mrna
DEFINITION	7h2a12.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:3316798 3',	EST
ACCESSION	BF002153	
VERSION	BF002153.1	GI:10702428
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
COMMENT	Tumor Gene Index	
	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-r@mail.nih.gov	
	Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,	
	Ph.D.	
	cDNA Library Preparation: M. Bento Soares, Ph.D.	
	cDNA Library Arrayed by: Greg Lennon, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	Clone distribution: NCI-CGAP clone distribution information can be	

found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 463.

FEATURES	
SOURCE	

```

1. 540
location/Quadrilateris
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cloned_into="IMAG:3316798"
/clone_id="NCLCGAP.Col6"
/tissue_type="colon tumor,
/lab_host="Drl10B"
/note="Organ: colon; Vector:
modified polylinker; Site:
plasmid DNA from the normal
prepared, and ss circles were
purification, this DNA was
hybridization reaction. The
from a pool of 5,000 clones,
(cloneIDs 1057416-1061255,
Subtraction by Bento Soares
93 c 85 q 1153
a

```

BASE C  
ORIGIN

Query Match 8.7%; Score 457; DB 143; Length 540;  
Best Local Similarity 99.8%; Pred. NO. 1.9e-189;  
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	4711	ctgcgctttaaacgacagtagtaccgcagctgttttatigagatctattgatattgaaagaatt	4770
Db	540	CTCGCGCTTTAAACGTACAGTACCAGCATGTTTTATTGGATCTATTGATTTGAAAAGAATTT	481
Qy	4771	tgttaggaLagatctttaagcagtaactctgcagtgttttgtattttgtattttctgcgaattt	4830
Db	480	TGTTTAGATAGATCTTAACGAGTAATCTGTCAGTGTTTTGATTTGTATTCTCTCAATTT	421
Qy	4831	tacttgtaaaaaaaatttgcctttcaacaatttgggtgctaatcttcttgatgctcaatttgt	4890
Db	420	TACTGTGAAAAAAAATTTGTTTTCACCAATTTGGTGTCATATTTCTTGATGTGCATATTTGT	361
Qy	4891	tggagaggtLaaatggctctctcccttttgcgtatcttaactagtgtttactctctgggaccc	4950
Db	360	TGGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACTAGTGTTTACTCTCGGGCACC	301
Qy	4951	cttaactcttcagaggtgctaaattgtctgccattacaccagaagagtgcccttgatagga	5010

RESULT 9  
AW303630/C

LOCUS	AW303630	542 bp	mrna	EST	18-JAN-2000
DEFINITION	xv21e03.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone IMAGE:2813788 3', mRNA sequence.				
ACCESSION	AW303630				
VERSION	AW303630.1 GI:6713319				



```

Qy 5047 ctggatactttacacactcagtagtattggtccacagaattttctgccccttcatgcga 5106
|||||
Db 208 CTTGATTACTTTACACCTCAGTAGTATTGTTGCCAGAAATTTTCTGGCCCTTCATGSCA 149
|||||
Qy 5107 atgaaaattttaagaagaagatttaagtagtattttaatttaaaagagtggtataaaat 5166
|||||
Db 148 ATGAAAATTTTAAGAAGAAGAGATTAAAGTAGATTAAATTTTAAAGAGAGTGTTTATAAAAT 89
|||||
Qy 5167 aatgactgaattctttatccattttatcattcttcattcttcagttttttatttaattctactgtat 5226
|||||
Db 88 AATGACTGAATCTTTATCCCAATTTTATCACTCTTTTATCACTCTTTTATTAATCTACTGTAT 29
|||||
Qy 5227 caataaaattctgtaattttaa 5248
|||||
Db 28 CAATAAAATCTGTAATTTGAA 7
|||||

RESULT 6
AI911784/c
LOCUS      AI911784      535 bp      mRNA      EST      17-DEC-1999
DEFINITION w32h01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
ACCESSION  AI911784
VERSION     AI911784.1 GI:5631639
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 535)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 708 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 446.
            Location/Qualifiers
                1..535
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2329873"
                /clone_lib="Soares_NFL_T_GBC_S1"
                /lab_host="DH10B"
                /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                a modified polylinker; Site.1: Not I; Site.2: Eco RI;
                Equal amounts of plasmid DNA from three normalized
                libraries (fetal lung NBHL19W, testis NHT, and B-cell
                NCI-CGAP_GCB1) were mixed, and ss circles were made in
                vitro. Following HAP purification, this DNA was used as
                tracer in a subtractive hybridization reaction. The driver
                was PCR-amplified cDNAs from pools of 5,000 clones made
                from the same 3 libraries. The pools consisted of
                I.M.A.G.E. clones 297480-302087, 682632-687239,
                726408-728711, and 729096-731399. Subtraction by Bento
                Soares and M. Fatima Bonaldo."
            BASE COUNT  205 a 90 c 81 g 159 t
            ORIGIN

Query Match      9.0%; Score 472; DB 103; Length 535;
Best Local Similarity 99.8%; Pred. No. 4.9e-196;
Matches 522; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4723 actacagatcccaactgtttatggatctattgaattgaagaagaattttagtagataga 4782
|||||
Db 535 ACTGACAGTACCCGACTGTTTATGGATCTATTGATTGAAGAAGAAATTTGTAGATAGA 476
|||||
Qy 4783 tcttaagcagtaactgtcagtggtttgtattgtatttctgcaatttttactgtgaaaaa 4842

```

```

Db 475 TCTTAAGCAGTAATCTGTGTCAGTGTGTTGATTGTTGATTCTGCAATTTTACTGTGAAAAA 416
|||||
Qy 4843 aatttttttcaacaatttggtgcattttcttgaatgacattttgttgagagattaaa 4902
|||||
Db 415 AAATTTGTTTCAACAATTTGGTGTGTCATTTCTTGATGTGTCATTTGTTGGAGAGTTAAA 356
|||||
Qy 4903 tggctcttccctttgtgtatcttaccatagttttactccctggccaccccttaattctcag 4962
|||||
Db 355 TGGTCTCTTCCTTTGTTGTCATCTTACCTAGTGTGTTACTCTGGCACCCCTTAATCTTCAG 296
|||||
Qy 4963 aggtgctaaaattgtctgcattacaccagaaggatgctctgtgataggaggacaaccatgc 5022
|||||
Db 295 AGGTGCTAAATTTGTCGCCATTACACAGAAGGATGCCCTCTGATAGGAGGACAACCATGC 236
|||||
Qy 5023 aattgtgaaatagtcctgaagttcttggtattcttaccactcagtagtattgtgtccc 5082
|||||
Db 235 AAATTTGGAATAGTCTCTGAAGTTCTTGGATTACTTTACACCTCAGTATTGATTGTCCCC 176
|||||
Qy 5083 agaattttctgcttcatgccaatgaaattttaagaagaagaatttaagtagtatttta 5142
|||||
Db 175 AGAATTTCTGCGCTTTCATGCGCAATGAAATTTTAAAGAAGAAGATTTAAAGTATTTTA 116
|||||
Qy 5143 attttaagagtggtgtttataaaataatgtagtactgaattctttatccattttatcatcctt 5202
|||||
Db 115 ATTTTAAAGAGCTGTTTATAAAATAATGCTACTGAATCTTTTATCCATTTTATCATCCTT 56
|||||
Qy 5203 tcaatttttataactactgtatcataataaaattctgtaattt 5245
|||||
Db 55 TCAGTTTTTATTAATCTACTCTATCAATAAAATTTCTGTAATTT 13
|||||

RESULT 7
AI655429/c
LOCUS      AI655429      480 bp      mRNA      EST      14-DEC-1999
DEFINITION ts98f06.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:2239331 3',
            mRNA sequence.
ACCESSION  AI655429
VERSION     AI655429.1 GI:4739408
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 480)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1048 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 455.
            Location/Qualifiers
                1..480
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2239331"
                /clone_lib="NCI-CGAP_GC6"
                /tissue_type="pooled germ cell tumors"
                /lab_host="DH10B"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Plasmid DNA from the normalized library

```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 648)  
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt  
J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
Quackenbush,J.

TITLE Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)

COMMENT The Institute for Genomic Research  
Contact: John Quackenbush  
7912 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 317  
Seq primer: Forward.

FEATURES Location/Qualifiers

source  
1..648  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE Resequences, MAGM"  
/note="Vector: pBluescriptSKm"

BASE COUNT 257 a 112 c 94 g 185 t

ORIGIN

Query Match 10.2%; Score 536; DB 122; Length 648;

Best Local Similarity 99.8%; Pred. No. 4e-224; Mismatches 1; Indels 0; Gaps 0;

Matches 586; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4665 tttttttatgtttctgattgtgaagtgctgtgtagctttatcttcggtgttaaac 4724

DB 587 TTTTTTTTAAATGTTCTGATTTCTGAAGTCTGTGTAGCTTTATCTGCGGCTTTAAAC 528

QY 4725 tgacagaccgactgtttatgtgattctattgattgaaagaattgttaggatgac 4784

DB 527 TGACAGTACCCGACTGTTTATTGGATCTATTGATTGAAAGAATTTGTAGGATAGATC 468

QY 4785 ttaagcagtaactgtcagtggttgattgttatcttcgcaatttactgtgaaaaaa 4844

DB 467 TTAAGCAGTAATCTGTCAGTGTGTGTTGTTGTTCTGCAATTTTACTGTGAAAAAA 408

QY 4845 attgttttcaacaattggtgttcattttcttcttgatgtcactattgttgagagtgaaatg 4904

DB 407 ATTGTGTTTCAACAATGGTGTCTATTTCTTGATGTCTACTATTGTTGGAGAGTTAAATG 348

QY 4905 gtctctccctttgtattcttaccctagttttactcctgggcaaccttaattcagag 4964

DB 347 GTCTCTCCCTTTGNGTATCTTACCTAGTGTGTTTACTCTGGGCACCTTTAACTTTCAGAG 288

QY 4965 gtgctaaattgtctgcattacaccagaagatgctcttgataggaggaacacacagcaa 5024

DB 287 GTGCTAAATGTCTGCATACACAGAGAGGATGCTCTGATAGGAGGACACCATGCCAA 228

QY 5025 attgtgaaatagtcctcgaagttcttgattactttacacctcagttattgttgcacag 5084

DB 227 ATTGTGAAATAGTCCCTGAAGTCTTGGATTACTTTTACACCTCAGTATTGATTGTCCACG 168

QY 5085 aattttctggccttccatggcaatgaaattttaagaagaagatttaagatttttaatt 5144

DB 167 AATTCTTGGCCCTTCATGGCAATGAAATTTTAAAGAAGAAGATTAAAGTATTTTAAAT 108

QY 5145 ttttaagagtggtgtataaaataatgtactgaaatttttattccattttatccatttc 5204

DB 107 TTTAAAGAGTGTGTTTAAATAATCTACTGAATCTTTTATCCATTTTATCATCTTTC 48

QY 5205 agtttttataactactgtatcataataaattctgtaatttgatga 5251

DB 47 AGTTTATTATTAATCTACTGTATCAATAAAAAATCTGTAATTGTAATGA 1

RESULT 5

AW629426/c

LOCUS

DEFINITION

hi57b12.x1 Soares NFL T\_GBC\_S1 Homo sapiens cDNA clone

IMAGE:2976383 3', mRNA sequence.

ACCESSION

AW629426

VERSION

AW629426.1

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 444.

FEATURES

source

1..568

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with

a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NbHL19W, Testis NHT, and B-cell

NCI-CGAP GC81) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 220 a 97 c 87 g 164 t

ORIGIN

Query Match

Best Local Similarity 9.7%; Score 511; DB 118; Length 568;

Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4687 ctgaagtgcttgatagcttttatctgcggctttaactgacagtagtaccgactgtttatt 4746

DB 568 CTGAAGTGCTTGATAGCTTTTATCTGCGGCTTTAAACTGACAGTACCCGACTGTTATT 509

QY 4747 ggatctatgattgaaaagaattttagatagatctttaagcagtaattctcagtggt 4806

DB 508 GGATCTATTGATTGAAAAAATTTGTTAGGATAGATCTTTAAGCAGTAAATCTGTCACTGT 449

QY 4807 ttgtattgtatttctgcgaattttactgtgaaaaaaattttttcaacaattggtgt 4866

DB 448 TTGTATTGTTATCTCTGCAATTTTACTGTGAAAAAAATTTGTTTCACAATTTGGTGT 389

QY 4867 cattttcttgatgctactatttgttgagagttaaatggtctcttccctttgtgtatctt 4926

DB 388 CATTTTCTTGATGTCATCTATTGTTGGAGAGTTAAATTTGTTCTTCCCTTTGTGTATCTT 329

QY 4927 acctagtgttactcctcggcacccttaattcttcagaggtgctaaattgtctccattac 4986

DB 328 ACCTAGTGTTTTACTCTCTGGCACCCCTTAATCTTTCAGAGGTGCTAAATTTGTTGCCATTAC 269

QY 4987 accagaagatgctctctgataggaggaacaccatgcaaatgtgaaatagctcctgaagt 5046

DB 268 ACCAGAGGATGCCTCTGTATAGGAGGACACCATGCAAAATTTGAAATAGTCTCTGAAGTT 209



ORIGIN

Query Match 11.3%; Score 596; DB 136; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.9e-250;  
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2752 atgaatcatcattagaaacattatcagctatgtcattagctatcaacgatgaatgct 2811  
|||||  
DB 1 ATGAAATCATCATATAGAACATATCAGCTATGTGCTATGCTATCAACGATGAATGCT 60  
|||||

QY 2812 atcaagtaagacaagtgtttgcccagagaaacttcacaaagccctttcccggttaccggttc 2871  
|||||  
DB 61 ATCAAGTAAGACAAGTGTTCGCCAGAACTTCACAAAGCCCTTCCCGTTTACGGCTTC 120  
|||||

QY 2872 cacttgagtatatgcaaatctgtgccccttggcgaagaagatcctgtaaaagagagaagag 2931  
|||||  
DB 121 CACTGTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAG 180  
|||||

QY 2932 ctcatgtaggcaagtgttggtaaaaaataataatgtaagcgaggagatctgaagcagc 2991  
|||||  
DB 181 CTGATGTAGGCAATGTTGGTGAAAAATATAATGTAAAGCGGAGATCTGAAGCAGC 240  
|||||

QY 2992 atgcagctgttagtgaataattattgtctctctaccagagatgttggttccatatcaaa 3051  
|||||  
DB 241 ATGCAGCTGTAGTGAATAATTATTGTCTCTTCTACAGAGATGTGTTCCATATACAA 300  
|||||

QY 3052 ttaacctttggcacatgacccagattatgtcaaaagacaggtatgtgaacacattaaag 3111  
|||||  
DB 301 TTCACTTTTGGCACATGACCCAGATTATGTCAAAAGTACAGGATATTGAACAATTAAAG 360  
|||||

QY 3112 atgttaagaagtcttgtgttctggaattattgctgaataattatgctgaataataaaca 3171  
|||||  
DB 361 ATGTTAAAGAAATGCTTTGTTGTTCTGGAATATTAAATGGCTTAAATGAATAATCA 420  
|||||

QY 3172 gtacgcgttttatcagaaagtgtagaaataatttaacaaacaaacaaagatgccaaaggac 3231  
|||||  
DB 421 GTCACGCTTTTATCAGAAAGATGTAGAAAAATATTAAACAAACAAAGATGCCCAAGGAC 480  
|||||

QY 3232 cagatgtgcaaaaataagaaataacatgtacactgtgtgtgtgtggtgcaatgaatca 3291  
|||||  
DB 481 CAGATGTATGCAAAATGAATGAATAACTGTACACTGTGTGTGTGTGTCATGAATATCA 540  
|||||

QY 3292 tcatgtcaaaagactacatacagatttgaattctcctaaagacccggtactacca 3347  
|||||  
DB 541 TCATGTCAAGAGTACTACATACAGTTTGGAACTCTCTAAAGACCCCGGTACTACCA 596  
|||||

RESULT 3  
BG256731  
LOCUS 734 bp mRNA EST 13-FEB-2001  
DEFINITION 602370939F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4479153 5', mRNA sequence.  
ACCESSION BG256731  
VERSION BG256731.1 GI:12766547  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: AFCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLAMI0310 row: m column: 10  
High quality sequence stop: 732.

FEATURES  
source Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4479153"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/site="Organ: testis; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 278 a 154 c 161 g 141 t  
ORIGIN

Query Match 10.9%; Score 576; DB 175; Length 734;  
Best Local Similarity 100.0%; Pred. No. 1.1e-241;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3318 ttggaatctctaaagaccgggtactaccagctgtttcttctactcaacctgcacaaagt 3377  
|||||  
DB 21 TTGGAATCTCTTAAGACCCGGTACTACCAGCTCGTTCTTCACTCAACCTGACAAGAT 80  
|||||

QY 3378 ttcagttaaccccaaaattatctgcctcctgaaatgaataatcttctactcctctggaaaa 3437  
|||||  
DB 81 TTCAGTAACACCAAAATTTATCTGCCCTCCTGAAATGAATCATTTTCACTCTCTGAAAA 140  
|||||

QY 3438 cctaaaaacaaatgtcttagagctgttaaacagccacttctcagcagcagaacaa 3497  
|||||  
DB 141 CCTAAAAACAACCAATGTTCTAGGAGCTGTTAAACAGGCCACTTTTCATCAGCAGGCAAGCAA 200  
|||||

QY 3498 tctcagaccataatcatcacgaatggaactgaaagcaatgcaagcagcagctcaaatcca 3557  
|||||  
DB 201 TCTCAGACCAAAATCATCAGCAATGGAATCTTAAGCAATGCAACGACGAGCTCAATCCA 260  
|||||

QY 3558 agctcctctggaagataaaagggaggctgtgattgtcttgaatctgcaatctgcaatctgcaat 3617  
|||||  
DB 261 AGCTCTCTCTGGAAGATAAAGGGGAGGCTTTGATAGTTCTGAAATGGATCAGAGTGAAT 320  
|||||

QY 3618 gaagattacaaatgtctcacccttgcgggggaaaaaagtgcagagagacgactct 3677  
|||||  
DB 321 GAAGATTACAAATGCTCTTCACCTTTGCCGGGAAAAAAGTGAACAGAGACGACTCT 380  
|||||

QY 3678 gatctgttaaggtctgaattggagaagccttagagcgaggagaaaaaacgcccgtcacagaa 3737  
|||||  
DB 381 GATCTCTTAAGGCTGAATTGGAGAAGCCTAGAGGCGAGGAAAAAACGCCCGTCCACAGAA 440  
|||||

QY 3738 caggagagaaaattagtgatggtatgactgaattggttacagggaacagaacacctaata 3797  
|||||  
DB 441 CAGGAGGAGAAATTTAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500  
|||||

QY 3798 ggcagtcagcgaactcgaaaaagagcgcatacgccttcagaatctgataacacagctgg 3857  
|||||  
DB 501 GGCAGTCAGCGAAGTCGGAAGAGAGGCCATACGGCTTCAGAACTCTGATGAACAGCAGTGG 560  
|||||

QY 3858 cctgagaaaaagaggtccaagaagatatattagaa 3893  
|||||  
DB 561 CCTGAGGAAAAAGAGGCTCAAGAGAGATATATTAGAA 596  
|||||

RESULT 4  
AW973437/c  
LOCUS 648 bp mRNA EST  
DEFINITION EST385430 MAGE resequences, MAGM Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW973437  
VERSION AW973437.1 GI:8164508  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens



REFERENCE 1 (bases 1 to 791)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbbrp/image/image.html  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES  
 source

1..791  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3086220"  
 /clone\_lib="NCI CGAP Sub8"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; NCI-CGAP-Sub8  
 is a subtracted library derived from NCI-CGAP-Sub5. The  
 NCI-CGAP-Sub8 library had 2.5 million recombinants. A  
 single-stranded DNA preparation of NCI-CGAP-Sub5 was used  
 as a tracer in a subtractive hybridization with a driver  
 comprising a pool of clones from NCI-CGAP-Sub5 (IMAGE  
 clone ids 2732833-2737415, 3068040-3069191; 25% of the  
 driver population), a pool of clones from NCI-CGAP-Sub4  
 (IMAGE clone ids 2723592-2729326; 25% of the driver  
 population), NCI-CGAP-Sub6 (pool AIF-AJU, IMAGE ids  
 2728969-2733190; 25% of the driver population), and  
 NCI-CGAP-Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550  
 ; 25% of the driver population). Subtraction was  
 performed as previously described (Bonaldo, Lennon &  
 Soares (1996): Normalization and Subtraction: Two  
 Approaches To Facilitate Gene Discovery. Genome Research  
 6: 791-806.  
 TAG\_LIB=NCI-CGAP\_Lu5  
 TAG\_TISSUE=Lung  
 TAG\_SEQ=CAAC"

BASE COUNT 270 a 151 c 144 g 226 t  
 ORIGIN

Query Match 12.0%; Score 631; DB 149; Length 791;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-266;  
 Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2745 ttttaccatgaatcattacattagaaataatcaactatgctgattagctatcaacgat 2804  
 DB 13 TGTACCAGTAATCATCATATAGAACCAATATACGCTATGTGCAATAGTATCAACGAT 72  
 QY 2805 gaatgctatcaagtaagacaagtgtttgccagaaacttcacaaagccctttccqttta 2864  
 DB 73 GAATGCTATCAGTAGACAGAGTGTGCCAGAACTTCACAAAGCCCTTCCCGTTTA 132  
 QY 2865 cggctccacttgatgataatgcaatctgtgccctttgtgcataaagatcctgttaaaggag 2924  
 DB 133 CGGCTTCCACTTGTAGTATATGCAATCTGTGCCCTTGTGTGCAAAAGATCCTGTAAAGGAG 192  
 QY 2925 aaaaagctcattaggaatgttttggaataataataatgtaagcgaggatctcg 2984  
 DB 193 AGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGCGGGAGTATCTG 252  
 QY 2985 aagcagatcagctgttagtgaataattattgtctctctaccagagtatgtgttcca 3044  
 DB 253 AGCAGCATGCAGCTGTAGTGAATAATATATGCTCTCTCTACCAGAGTATGTTGTCCA 312

QY 3045 tatacaattcaacttttggcacatgaccagattatgtcaaaatcacagattattgaacaa 3104  
 DB 313 TATACAATTCACCTTTTGGCACATGACCAGATATATGTCAAAGTCAGAGTATTTGAACA 372  
 QY 3105 cttaaagatgtaaaagaatgcttttggtttttcttgaaatatttaataatgctaaatgaa 3164  
 DB 373 CTTAAAGATGTTAAAGAATGCTCTTGGTTTCTGGAATATTAATGCTGCTAAATGAA 432  
 QY 3165 aataacagtcacgcttttattcagaagaatggttagaataattatacaacaaagaatgccc 3224  
 DB 433 AATAACAGTCACGCTTTTATCAGAAAGATGTTAGTAATAATATTAACAAACAAAGATGCC 492  
 QY 3225 caaggaccagatgatgcaaaaatgaatgaaaactgtacactgtgtgtatgttgcacatg 3284  
 DB 493 CAAGGACCAGATGATGCAAAAATGAATGAAAACTGTACACTGTGTGTATGTTGCCATG 552  
 QY 3285 aatatcatcatgcaagaatgactacatcacagtttggatctctctaaagaccggtacta 3344  
 DB 553 AATATCATCATGTCAAAGAGTACTACATACAGTTTGAATCTCTTAAGACCCGGTACTA 612  
 QY 3345 ccagctcgtttcttctcaactcaacctgacaaga 3375  
 DB 613 CCAGCTCGTTCTTCTCACTCAACTGACAGA 643  
 RESULT 2  
 BE504550  
 LOCUS  
 DEFINITION  
 BE504550 596 bp mRNA EST 04-AUG-2000  
 hz58a05.x1 NCI-CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3212144.3  
 similar to TR:094237 O94237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA  
 sequence.  
 ACCESSION BE504550  
 VERSION BE504550.1 GI:9706958  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 596)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 458.  
 Location/Qualifiers  
 1..596  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3212144"  
 /clone\_lib="NCI-CGAP\_Lu24"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Plasmid DNA from the normalized  
 library NCI-CGAP\_Lu5 was prepared, and ss circles were  
 made in vitro. Following HAP purification, this DNA was  
 used as tracer in a subtractive hybridization reaction.  
 The driver was PCR-amplified cDNAs from a pool of 5,000  
 clones made from the same library (cloneIDs  
 1414920-1417991 and 1520904-1522439). Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 211 a 103 c 117 g 165 t

BASE COUNT



871	26	0.5	245	105	AL366935	AL366935 MtBaliB04	944	26	0.5	285	132	BB356786	BB356786	
872	26	0.5	246	20	AI447785	AI447785 mq84e01.x	c 945	26	0.5	286	27	AV258498	AV258498	
c 873	26	0.5	247	28	AV282689	AV282689	946	26	0.5	287	118	AW633095	AW633095	
c 874	26	0.5	248	118	AW638728	AW638728 bl73f04.w	c 947	26	0.5	288	6	AA358545	AA358545	
c 875	26	0.5	249	188	T06477	T06477 EST04366.Fe	c 948	26	0.5	289	27	AV269715	AV269715	
c 876	26	0.5	250	119	AW641529	AW641529 cm08d09.w	c 949	26	0.5	290	118	AW633122	AW633122	
c 877	26	0.5	251	249	BF991574	BF991574 MR3-GN014	c 950	26	0.5	291	288	AQ093568	HS_3026_B	
c 878	26	0.5	252	249	A2846593	A2846593 2M0146C23	c 951	26	0.5	292	288	AZ346917	LM0082F19	
c 879	26	0.5	253	4	AA237598	AA237598 mx11h04.r	c 952	26	0.5	293	114	AW305335	AW305335	
c 880	26	0.5	254	22	AT161212	AT161212 tt58b12.x	c 953	26	0.5	294	256	B33481	B33481	
c 881	26	0.5	255	111	AW088375	AW088375 x09e08.x	c 954	26	0.5	295	164	BE180440	RC3-HW062	
c 882	26	0.5	256	139	BE721605	BE721605 189191.MA	c 955	26	0.5	296	103	AI925988	AW38h12.x	
c 883	26	0.5	257	118	AW632874	AW632874 bl01c10.x	c 956	26	0.5	297	118	AW632841	bl01a01.x	
884	26	0.5	258	244	AZ494100	AZ494100 LM0329H10	c 957	26	0.5	298	9	AA584275	AA584275	
c 885	26	0.5	259	19	AT1335296	AT1335296 ta47b03.x	c 958	26	0.5	299	118	AW635098	bl29c05.s	
c 886	26	0.5	260	118	AW633139	AW633139 bl04c01.x	c 959	26	0.5	300	4	AA277233	va79h12.r	
c 887	26	0.5	261	188	R99438	R99438 YQ79e04.r1	c 960	26	0.5	301	118	AW633029	bl03a06.x	
c 888	26	0.5	262	19	AI392442	AI392442 NCSC3C12T	c 961	26	0.5	302	4	AA260406	va95e08.r	
c 889	26	0.5	263	22	AI612027	AI612027 tt63d07.x	c 962	26	0.5	303	4	AA265180	mx90d11.r	
c 890	26	0.5	264	22	AI621496	AI621496 486092B01	c 963	26	0.5	304	110	AV739720	AV739720	
c 891	26	0.5	265	27	AV264213	AV264213 AV264213	c 964	26	0.5	305	110	AV739720	AV739720	
c 892	26	0.5	266	113	AW632396	AW632396 qx88f02.x	c 965	26	0.5	306	149	BF478508	WHE2009.E	
c 893	26	0.5	267	241	AZ311674	AZ311674 LM0027010	c 966	26	0.5	307	21	AI523898	ot66c04.s	
c 894	26	0.5	268	3	AA212432	AA212432 mw76c09.r	c 967	26	0.5	308	109	AV044330	AV044330	
c 895	26	0.5	269	22	AI580595	AI580595 tb71q06.x	c 968	26	0.5	309	136	BE488154	267700.Am	
c 896	26	0.5	270	102	AT862257	AT862257 tb90a02.x	c 969	26	0.5	310	118	AW634586	bl12e05.w	
c 897	26	0.5	271	150	BF560275	BF560275 UI-R-El-g	c 970	26	0.5	311	156	C22967	C22967	
c 898	26	0.5	272	165	BE234741	BE234741 141958.MA	c 971	26	0.5	312	5	AA299178	EST11663	
c 899	26	0.5	273	107	AW037846	AW037846 AU037846	c 972	26	0.5	313	110	AV740764	AV740764	
c 900	26	0.5	274	118	AW632889	AW632889 bl01e01.x	c 973	26	0.5	314	116	AW449415	UI-H-B13-	
c 901	26	0.5	275	118	AW632938	AW632938 bl02a03.x	c 974	26	0.5	315	118	AW633201	bl04h04.x	
c 902	26	0.5	276	19	AI344828	AI344828 ta99f10.x	c 975	26	0.5	316	170	BF853708	MR2-EN009	
c 903	26	0.5	277	118	AW637298	AW637298 bl57a08.w	c 976	26	0.5	317	251	AZ883405	RPTI-23-1	
c 904	26	0.5	278	171	BF930587	BF930587 MR2-NT013	c 977	26	0.5	318	118	AW633096	bl03g04.x	
c 905	26	0.5	279	20	AI462830	AI462830 vb49f10.x	c 978	26	0.5	319	155	BG608477	307077.MA	
c 906	26	0.5	280	144	BF077836	BF077836 227994.MA	c 979	26	0.5	320	170	BF874487	RC1-ET013	
c 907	26	0.5	281	118	AW633097	AW633097 bl03g05.x	c 980	26	0.5	321	300	173	BG087728	H3143A07-
c 908	26	0.5	282	118	AW633104	AW633104 bl03h01.x	c 981	26	0.5	322	118	AW633135	bl04b09.x	
c 909	26	0.5	283	121	AW834185	AW834185 MR2-TT001	c 982	26	0.5	323	118	AW633141	bl04c03.x	
c 910	26	0.5	284	165	BE235913	BE235913 143540.MA	c 983	26	0.5	324	162	BB605008	BB605008	
c 911	26	0.5	285	110	AV738603	AV738603 AV738603	c 984	26	0.5	325	302	238	AZ108150	RPTI-23-4
c 912	26	0.5	286	113	AW264971	AW264971 qx58f10.x	c 985	26	0.5	326	18	AI267659	aq92h02.x	
c 913	26	0.5	287	9	AA623752	AA623752 vg69b04.s	c 986	26	0.5	327	33	AV225348	AV225348	
c 914	26	0.5	288	7	AA469262	AA469262 nc69f04.s	c 987	26	0.5	328	111	AW085910	xc75c10.x	
c 915	26	0.5	289	28	AV278690	AV278690	c 988	26	0.5	329	121	AW868509	MR1-SN006	
c 916	26	0.5	290	27	AV277250	AV277250	c 989	26	0.5	330	143	BF001154	7q59a05.x	
c 917	26	0.5	291	249	AT773912	AT773912 2M0001F13	c 990	26	0.5	331	160	BB520204	BB520204	
c 918	26	0.5	292	22	AI610952	AI610952 tt60f11.x	c 991	26	0.5	332	4	AA244626	mx07e11.r	
c 919	26	0.5	293	112	AW185154	AW185154 se87e10.y	c 992	26	0.5	333	7	AA426459	zW02a03.r	
c 920	26	0.5	294	240	AZ242469	AZ242469 RPTI-23-7	c 993	26	0.5	334	27	AV258107	AV258107	
c 921	26	0.5	295	118	AW633055	AW633055 bl03c08.x	c 994	26	0.5	335	304	151	BF592897	7j196d09.x
c 922	26	0.5	296	118	AW633146	AW633146 bl04c08.x	c 995	26	0.5	336	162	BE014014	BE014014	
c 923	26	0.5	297	169	BF774517	BF774517 mx20177.MA	c 996	26	0.5	337	3	AA208851	mw75803.r	
c 924	26	0.5	298	4	AA242620	AA242620 tm08d06.r	c 997	26	0.5	338	122	AW971300	AW971300	
c 925	26	0.5	299	111	AW104134	AW104134 xd65e03.x	c 998	26	0.5	339	3	AA212774	zk25d06.s	
c 926	26	0.5	300	125	BB072629	BB072629 BB072629	c 999	26	0.5	340	166	BE321246	mw78g06.r	
c 927	26	0.5	301	22	AI582986	AI582986 LI98a03.x	1000	26	0.5					
c 928	26	0.5	302	164	BE148687	BE148687 MR0-HT024								
c 929	26	0.5	303	27	AV265901	AV265901								
c 930	26	0.5	304	118	AW633115	AW633115 bl03h12.x								
c 931	26	0.5	305	140	BE819309	BE819309 CMI-BN032								
c 932	26	0.5	306	247	AZ659664	AZ659664 LM0537M07								
c 933	26	0.5	307	18	AI287290	AI287290 qv60f02.x								
c 934	26	0.5	308	118	AW633036	AW633036 bl03b01.x								
c 935	26	0.5	309	281	AW633036	AW633036 bl03b01.x								
c 936	26	0.5	310	119	AW682376	AW682376 EST01185								
c 937	26	0.5	311	13	AA312454	AA312454 cm52b04.s								
c 938	26	0.5	312	19	AI349675	AI349675 tb58d06.x								
c 939	26	0.5	313	114	AW301766	AW301766 xr83d06.x								
c 940	26	0.5	314	118	AW633127	AW633127 bl04a12.x								
c 941	26	0.5	315	120	AW803530	AW803530 IL2-UM008								
c 942	26	0.5	316	230	AQ566683	AQ566683 HS_2104.B								
c 943	26	0.5	317	249	AZ781349	AZ781349 2M0019A09								
c 944	26	0.5	318	110	AV739714	AV739714 AV739714								

ALIGNMENTS

RESULT 1

BF509252

LOCUS

DEFINITION

UI-H-B14-aow-c-07-0-ui.sl NCI\_CGAP\_Sub8 Homo sapiens cDNA clone IMAGE:3086220 3', mRNA sequence.

ACCESSION

BF509252

VERSION

BF509252.1

KEYWORDS

GI:11592550

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## ALIGNMENTS

## RESULT 1

BF509252	791 bp	mRNA	EST	06-DEC-2000
UI-H-B14-aov-c-07-0-UI.s1	NCI_CGAP_Sub8	Homo sapiens	cdna	clone
IMAGE:3086220 3',	mRNA	sequence.		
BF509252	GI:11592550			
BF509252	EST.			
BF509252	KEYWORDS			
BF509252	SOURCE			
BF509252	ORGANISM			
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				



c 725	125	119	AW642236	AW642236 cml7d03.w	798	26	0.5	202	241	AZ331604	AZ331604	1M0059H07
c 726	131	18	AI310949	AI310949 ta76d08.x	799	26	0.5	203	164	BE150909	BE150909	RC4-HF027
c 727	133	118	AW639663	AW639663 bl85e10.w	c 800	26	0.5	204	22	AI588437	AI588437	fc02a06.x
c 728	136	19	AI345273	AI345273 tb68a10.x	c 801	26	0.5	204	118	AW633071	AW633071	bl13e02.x
c 729	139	118	AW633102	AW633102 bl03g10.x	c 802	26	0.5	204	118	AW633756	AW633756	bl11g05.w
c 730	144	18	AI310875	AI310875 ta87a05.x	c 803	26	0.5	204	118	AW633766	AW633766	bl11h04.w
c 731	144	190	W39837	W39837 333 Mouse v	c 804	26	0.5	205	112	AW185166	AW185166	se87g03.y
c 732	145	165	BE235153	BE235153 142506 MA	c 805	26	0.5	205	118	AW633066	AW633066	bl03d08.x
c 733	146	3	AA209567	AA209567 mw75a07.r	c 806	26	0.5	206	103	AI872843	AI872843	wm71f03.x
c 734	146	137	AI371638	AI371638 tb72h09.x	c 807	26	0.5	206	118	AW633165	AW633165	bl04e03.x
c 735	146	122	AW946864	AW946864 RC2-EF002	c 808	26	0.5	206	170	BF890592	BF890592	CM1-MF010
c 736	147	118	AW632958	AW632958 bl02c01.x	c 809	26	0.5	208	19	AI340523	AI340523	tb30c07.x
c 737	148	3	AA207069	AA207069 zr87e06.s	c 810	26	0.5	209	20	AI342756	AI342756	th49b03.x
c 738	148	18	AI249411	AI249411 qx74g02.x	c 811	26	0.5	209	118	AW633025	AW633025	bl03a01.x
c 739	149	137	BE610701	BE610701 sg64g11.y	c 812	26	0.5	209	164	BE151756	BE151756	QV0-HF030
c 740	150	17	AI232458	AI232458 EST229h146	c 813	26	0.5	210	22	AI613386	AI613386	tl80e12.x
c 741	150	19	AI345528	AI345528 tb69h03.x	c 814	26	0.5	211	18	AI270927	AI270927	qw52b09.x
c 742	151	118	AW633195	AW633195 bl04g10.x	c 815	26	0.5	211	171	BF909224	BF909224	PM3-UT005
c 743	154	118	AW633164	AW633164 bl04e02.x	c 816	26	0.5	212	22	AI630676	AI630676	ad14c10.x
c 744	154	137	BE548041	BE548041 601072072	c 817	26	0.5	212	118	AW633169	AW633169	bl04e08.x
c 745	155	17	AI223731	AI223731 qx32b03.x	c 818	26	0.5	212	132	BB371564	BB371564	BB371564
c 746	157	162	BE062038	BE062038 RC1-BF025	c 819	26	0.5	213	4	AA237192	AA237192	mx17h03.r
c 747	158	8	AA514673	AA514673 nf59e05.s	c 820	26	0.5	213	19	AI377146	AI377146	tb89e08.x
c 748	159	165	BE234108	BE234108 140717 MA	c 821	26	0.5	213	118	AW633186	AW633186	bl04g01.x
c 749	160	7	AA417829	AA417829 zv01a10.r	c 822	26	0.5	214	18	AI273880	AI273880	qu09d01.x
c 750	162	175	BG237164	BG237164 sab04d09.y	c 823	26	0.5	214	170	BF877366	BF877366	QV2-EF010
c 751	163	103	AI889818	AI889818 wm63f07.x	c 824	26	0.5	215	19	AI401159	AI401159	tg26f01.x
c 752	163	174	BG166238	BG166238 602345407	c 825	26	0.5	215	120	AW95810	AW95810	MR4-UM002
c 753	164	1	AA059132	AA059132 zf64c09.s	c 826	26	0.5	219	22	AI630306	AI630306	ad08c12.y
c 754	164	107	AU074500	AU074500 AU074500	c 827	26	0.5	219	110	AV739377	AV739377	AV739377
c 755	164	122	AW891165	AW891165 PM2-NF007	c 828	26	0.5	220	4	AA244598	AA244598	mx31n06.r
c 756	165	102	AI792457	AI792457 q168a09.y	c 829	26	0.5	221	104	AI990172	AI990172	ws38b01.x
c 757	165	111	AW501954	AW501954 UI-HF-BN0	c 830	26	0.5	221	114	AW301935	AW301935	xf85b01.x
c 758	167	67	DMAA0479	AA40479 LD12542.5	c 831	26	0.5	222	19	AI373773	AI373773	qz54f05.x
c 759	169	137	BE552916	BE552916 946087D01	c 832	26	0.5	223	18	AI310027	AI310027	tb49b10.x
c 760	171	102	AI799369	AI799369 tw33e10.x	c 833	26	0.5	223	121	AW856291	AW856291	RC1-CT028
c 761	171	130	BB277223	BB277223 BB277223	c 834	26	0.5	223	151	BF591791	BF591791	7150e05.x
c 762	172	7	AA30552	AA30552 zw22g06.s	c 835	26	0.5	223	155	BG610615	BG610615	727915 MA
c 763	172	118	AW633003	AW633003 bl02g03.x	c 836	26	0.5	224	17	AI242020	AI242020	qv51f10.x
c 764	175	18	AI311636	AI311636 ta43h10.x	c 837	26	0.5	224	22	AI580598	AI580598	tb71g09.x
c 765	175	103	AI906649	AI906649 RC-BT122-	c 838	26	0.5	224	23	AI675727	AI675727	wc40g04.x
c 766	175	169	BF814530	BF814530 MR2-CJ018	c 839	26	0.5	225	12	AA809479	AA809479	ob85c02.s
c 767	176	110	AV740432	AV740432 AV740432	c 840	26	0.5	225	122	AW897965	AW897965	RC3-NN006
c 768	176	110	AV741164	AV741164 AV741164	c 841	26	0.5	227	22	AI580614	AI580614	tb72a05.x
c 769	176	110	AV741894	AV741894 AV741894	c 842	26	0.5	230	170	BF875604	BF875604	QV3-EF010
c 770	176	118	AW638291	AW638291 bl68c12.w	c 843	26	0.5	231	24	AI758760	AI758760	ty15h06.x
c 771	177	142	BE964812	BE964812 601658467	c 844	26	0.5	231	139	BE693049	BE693049	SAC503 MO
c 772	179	10	AA691308	AA691308 vs14e09.r	c 845	26	0.5	232	118	AW633106	AW633106	bl03h03.x
c 773	180	102	AI792533	AI792533 q177a12.y	c 846	26	0.5	233	114	AW301411	AW301411	x876f01.x
c 774	180	118	AW640434	AW640434 bl195805.w	c 847	26	0.5	233	120	AW748972	AW748972	RC1-BF031
c 775	181	22	AI612793	AI612793 tp44f02.x	c 848	26	0.5	233	172	BG004526	BG004526	IL0-GN028
c 776	183	18	AI284741	AI284741 tq24e01.x	c 849	26	0.5	233	240	AZ267979	AZ267979	RPCI-23-1
c 777	184	19	AI349605	AI349605 ta73g12.x	c 850	26	0.5	234	123	BB018787	BB018787	BB018787
c 778	185	18	AI289689	AI289689 gw42e10.x	c 851	26	0.5	235	3	AA208776	AA208776	mw73f02.r
c 779	185	3	AA212578	AA212578 mw77f09.r	c 852	26	0.5	235	18	AI306755	AI306755	qw44b05.x
c 780	187	118	AW633150	AW633150 bl04c12.x	c 853	26	0.5	235	19	AI369117	AI369117	qw43d05.x
c 781	187	122	AW892681	AW892681 CM3-NN000	c 854	26	0.5	235	20	AI436486	AI436486	tb51f10.x
c 782	189	118	AW632950	AW632950 bl02b03.x	c 855	26	0.5	236	19	AI349974	AI349974	th59a05.x
c 783	189	118	AW633207	AW633207 bl04h11.x	c 856	26	0.5	236	22	AI567673	AI567673	tr64b02.x
c 784	191	164	BE160102	BE160102 QV1-HF041	c 857	26	0.5	236	123	BB016612	BB016612	BR016612
c 785	192	151	BF654195	BF654195 278266 MA	c 858	26	0.5	238	4	AA244555	AA244555	mx04h11.r
c 786	193	118	AW633185	AW633185 bl04f12.x	c 859	26	0.5	239	118	AW636849	AW636849	av51d08.w
c 787	193	163	BE138942	BE138942 xr63c12.x	c 860	26	0.5	240	28	AV312773	AV312773	BS121773
c 788	194	12	AA822798	AA822798 vq91b11.r	c 861	26	0.5	240	103	AI880491	AI880491	ac78h07.x
c 789	195	118	AW633000	AW633000 bl02f12.x	c 862	26	0.5	240	107	AU073621	AU073621	AU073621
c 790	196	175	BG298023	BG298023 602396236	c 863	26	0.5	241	21	AI537464	AI537464	tp04b11.x
c 791	199	19	AI343135	AI343135 tb04f10.x	c 864	26	0.5	241	27	AV270039	AV270039	qv66g12.x
c 792	200	115	AW426270	AW426270 60019 MAR	c 865	26	0.5	242	18	AI275677	AI275677	qv66g12.x
c 793	200	116	AW484318	AW484318 60018 MAR	c 866	26	0.5	243	118	AW633159	AW633159	bl04d09.x
c 794	201	118	AW633208	AW633208 bl04h12.x	c 867	26	0.5	243	172	BF989949	BF989949	MR2-GN012
c 795	201	122	AW905296	AW905296 QV2-NN107	c 868	26	0.5	244	118	AW632844	AW632844	bl01a04.x
c 796	202	18	AI269268	AI269268 q125e04.x	c 869	26	0.5	245	19	AI334840	AI334840	qw43e01.x
c 797	202	111	AW071272	AW071272 x863c06.x	c 870	26	0.5	245	21	AI521049	AI521049	to70c04.x



579	27	0.5	514	20	A1436455	ti09c11.x	C 652	27	0.5	669	247	A2648112	1M0517A05
580	27	0.5	517	237	A2072300	RPCI-23-3	C 653	27	0.5	670	248	A2738195	RPCI-24-1
581	27	0.5	520	242	A2339006	1M0070L13	C 654	27	0.5	677	103	A1893653	mn18b05.x
582	27	0.5	521	242	A2368976	1M0119M17	C 655	27	0.5	686	241	A2324179	1M0046F03
583	27	0.5	525	14	A1011151	EST205602	C 656	27	0.5	693	233	AQ742061	HS_5570_B
584	27	0.5	526	18	A1287470	qv81e09.x	C 657	27	0.5	728	236	AQ988648	RPCI-23-3
585	27	0.5	528	110	AV754096	AV754096	C 658	27	0.5	738	168	BF678033	602085072
586	27	0.5	532	11	AA744512	ny65c07.s	C 659	27	0.5	744	168	BF681835	602117674
587	27	0.5	533	110	AV755215	AV755215	C 660	27	0.5	745	233	AQ749077	HS_5538_A
588	27	0.5	533	233	AQ781023	HS_3169_B	C 661	27	0.5	748	22	AI567787	ti62f01.x
589	27	0.5	535	221	CNS04FGJ	Tetraodon	C 662	27	0.5	751	32	AV710961	AV710961
590	27	0.5	535	242	A2353018	A2353018	C 663	27	0.5	756	107	AU123336	AU123336
591	27	0.5	536	16	A1159622	uf02f05.x	C 664	27	0.5	765	146	BF245407	BF245407
592	27	0.5	538	223	AQ043211	CIT-HSP-2	C 665	27	0.5	799	169	BF792701	602253895
593	27	0.5	543	117	AM538870	C0112H04-	C 666	27	0.5	801	22	AI624300	ts28H08.x
594	27	0.5	546	167	BE430276	BE430276	C 667	27	0.5	805	114	AW332494	S9B AGS-
595	27	0.5	547	23	A1651924	w51b10.x	C 668	27	0.5	805	146	BF244960	601864371
596	27	0.5	547	246	A2563047	RPCI-23-2	C 669	27	0.5	822	219	CNS009VO	AI053991
597	27	0.5	548	111	AM070218	xa09d02.x	C 670	27	0.5	827	123	AW983107	HVSMEG000
598	27	0.5	548	156	C99229	C99229	C 671	27	0.5	839	174	BG167235	602345386
599	27	0.5	548	164	BE186233	945038A02	C 672	27	0.5	845	144	BF107623	601823981
600	27	0.5	549	24	AI742899	wg43b03.x	C 673	27	0.5	853	235	AQ894199	HS_3133_A
601	27	0.5	549	121	AW840682	RC4-CN000	C 674	27	0.5	867	168	BF678635	602085214
602	27	0.5	550	21	AI503368	vn87c08.x	C 675	27	0.5	872	233	AQ746568	HS_2278_A
603	27	0.5	553	118	AW634145	bl18b04.w	C 676	27	0.5	879	80	BF271208	GA_Eb001
604	27	0.5	555	22	AI567549	tr89g08.x	C 677	27	0.5	886	154	BG501720	602548777
605	27	0.5	555	172	BF969857	602272436	C 678	27	0.5	895	233	AQ742145	HS_5383_B
606	27	0.5	556	251	A2913961	RPCI-24-1	C 679	27	0.5	921	233	AQ746591	HS_2278_A
607	27	0.5	558	256	BE5614	CIT-HSP-202	C 680	27	0.5	966	222	CNS0522L	Tetraodon
608	27	0.5	562	112	AW184368	fj13b09.y	C 681	27	0.5	966	222	CNS05FAY	AI334771
609	27	0.5	567	9	AA599483	aq23f01.s	C 682	27	0.5	1369	143	BF038819	Tetraodon
610	27	0.5	568	136	BE473339	sp51a10.y	C 683	27	0.5	49	4	AA254893	601462003
611	27	0.5	570	32	AV717482	AV717482	C 684	26	0.5	49	4	AA254893	601462003
612	27	0.5	570	241	A2311236	1M0026M02	C 685	26	0.5	49	4	AA254893	601462003
613	27	0.5	575	118	AW637920	bl64a11.w	C 686	26	0.5	53	118	AW632856	6010b04.x
614	27	0.5	579	114	AW331093	707047B04	C 687	26	0.5	56	118	AW632959	6010b04.x
615	27	0.5	580	224	AQ078766	CIT-HSP-2	C 688	26	0.5	62	18	AI252328	qv27b08.x
616	27	0.5	583	241	A2281083	RPCI-23-1	C 689	26	0.5	62	118	AW633014	bl02b02.x
617	27	0.5	585	248	A2736910	RPCI-24-7	C 690	26	0.5	67	22	AI612006	tl63b03.x
618	27	0.5	585	250	A2843882	2M0142M22	C 691	26	0.5	67	118	AW633089	bl03f09.x
619	27	0.5	592	122	AQ349378	RPCI11-11	C 692	26	0.5	71	118	AW632877	bl01d01.x
620	27	0.5	593	112	AW192764	xl50h05.x	C 693	26	0.5	72	18	AI254432	qv34d05.x
621	27	0.5	593	229	AQ488486	RPCI-11-2	C 694	26	0.5	73	118	AW633155	bl04d05.x
622	27	0.5	594	235	AQ899648	HS-2046_B	C 695	26	0.5	75	22	AI582951	tr98e02.x
623	27	0.5	597	244	A2468128	1M0281A05	C 696	26	0.5	76	18	AI250613	qx33h09.x
624	27	0.5	602	144	BF131953	601820931	C 697	26	0.5	76	18	AI250615	qx33h11.x
625	27	0.5	605	241	A2333993	1M0063N07	C 698	26	0.5	77	168	BF698362	602130528
626	27	0.5	605	248	A2741812	RPCI-24-1	C 699	26	0.5	80	118	AW632978	bl02d10.x
627	27	0.5	606	242	A2378690	1M0133L04	C 700	26	0.5	84	12	AA824083	vr66b12.s
628	27	0.5	606	243	A2395910	1M0160P15	C 701	26	0.5	86	118	AW632878	bl01d02.x
629	27	0.5	607	241	A2313543	1M0029P12	C 702	26	0.5	90	18	AI254562	qv49a11.x
630	27	0.5	607	249	A2793664	2M0047L01	C 703	26	0.5	90	18	AI254572	qv49b11.x
631	27	0.5	612	238	AX121188	RPCI-23-1	C 704	26	0.5	92	4	AA253693	mx85c01.r
632	27	0.5	617	105	AL120173	DKE2P761C	C 705	26	0.5	92	4	AA253693	mx85c01.r
633	27	0.5	618	251	A2874453	2M0188002	C 706	26	0.5	93	19	AI340780	tb54a10.x
634	27	0.5	627	230	AQ552075	RPCI-11-4	C 707	26	0.5	93	249	A2778426	2M0013A12
635	27	0.5	628	244	A2455001	1M0257M10	C 708	26	0.5	95	104	AI945159	bs09f06.y
636	27	0.5	631	119	AW727071	GA_Ea001	C 709	26	0.5	99	19	AI376470	tb76e02.x
637	27	0.5	631	226	AQ284603	RPCI11-78	C 710	26	0.5	99	111	AW117089	fei8e02.x
638	27	0.5	632	242	A2376810	1M0130P18	C 711	26	0.5	101	244	A2453943	1M0255J08
639	27	0.5	633	112	AW180978	MgA0129r	C 712	26	0.5	107	119	AW681536	EST00254
640	27	0.5	637	238	A2107481	RPCI-23-3	C 713	26	0.5	109	118	AW632990	bl02e12.x
641	27	0.5	639	20	AI445562	tj30g02.x	C 714	26	0.5	111	22	AI581203	tb70e03.x
642	27	0.5	641	144	BF131261	601819517	C 715	26	0.5	111	22	AI612127	tl58e02.x
643	27	0.5	643	242	A2376123	1M0129P18	C 716	26	0.5	112	146	BF242774	601877651
644	27	0.5	644	241	A2336755	1M0067J08	C 717	26	0.5	112	161	BB585517	BB585517
645	27	0.5	648	249	A2795850	2M0051J17	C 718	26	0.5	114	114	AW346484	28203 MAR
646	27	0.5	649	103	AI1889305	wm97c10.x	C 719	26	0.5	116	244	A2494173	1M0329N16
647	27	0.5	649	168	BF675817	602083736	C 720	26	0.5	117	18	AI254661	qv57f01.x
648	27	0.5	651	142	BE964060	601657538	C 721	26	0.5	121	18	AI310902	ta87c10.x
649	27	0.5	653	175	BE283218	602406910	C 722	26	0.5	122	19	AI349769	ta87c10.x
650	27	0.5	667	220	CNS027BZ	Tetraodon	C 723	26	0.5	123	19	AI349299	ta75f12.x
651	27	0.5	668	243	A2422289	1M0200I11	C 724	26	0.5	125	118	AW576176	UI-HF-BNO



433	315	10	AA701131	AA701131 z956d10.s	c	506	27	0.5	415	104	AJ284206	AJ284206 4A3B-AAU-
434	318	249	AZ784921	AZ784921 2M0028K14	507	27	0.5	419	19	BE333107	BE333107 qq26h08.x	
435	319	BE070216	BE070216 QVA-BT040	508	27	0.5	420	136	BE502535	BE502535 hy16d04.x		
436	321	125	BB093363	BB093363 BB093363	509	27	0.5	420	187	AA088951	AA088951 mm51h08.r	
437	324	120	AA794731	AA794731 RC6-UM001	c	510	27	0.5	421	2	AA088951	AA088951 mm51h08.r
438	326	238	AZ093063	AZ093063 RPCI-23-4	c	511	27	0.5	423	249	AZ778114	AZ778114 2M0013E12
439	327	20	AI451720	AI451720 mx26c09.x	c	512	27	0.5	425	17	AJ236839	AJ236839 EST33401
440	331	26	AV209171	AV209171 AV209171	c	513	27	0.5	425	23	AI656555	AI656555 ts51f05.x
441	331	105	AI133537	AI133537 DKF2p762E	c	514	27	0.5	425	168	BF690990	BF690990 602247022
442	332	22	AI564682	AI564682 tq78q08.x	c	515	27	0.5	426	141	BE889073	BE889073 601513512
443	332	BB159288	BB159288 BB159288	516	27	0.5	428	23	AI651930	AI651930 wb51c04.x		
444	333	126	BB114478	BB114478 BB114478	517	27	0.5	430	15	AI051289	AI051289 ox24b02.x	
445	334	162	BE032563	BE032563 131966.MA	518	27	0.5	430	23	AI650430	AI650430 wa90g01.x	
446	334	175	BG233438	BG233438 dae28b11.	c	519	27	0.5	431	142	BE968824	BE968824 601649848
447	338	22	AI580655	AI580655 to72a09.x	c	520	27	0.5	433	104	AL022903	AL022903 t8130b51
448	339	167	BE416321	BE416321 MUG007.E0	c	521	27	0.5	438	20	AI451807	AI451807 mx10e11.y
449	340	165	BE232022	BE232022 136928.MA	c	522	27	0.5	438	104	AI954690	AI954690 wq32a07.y
450	340	172	BF986207	BF986207 RC3-GN004	c	523	27	0.5	438	234	AQ820872	AQ820872 HS_5263_B
451	342	223	AQ058580	AQ058580 CIR-HSP-2	524	27	0.5	439	118	AW632847	AW632847 b101a07.x	
452	343	118	AW634283	AW634283 b118906.w	525	27	0.5	441	22	AI582556	AI582556 ts02e11.x	
453	345	244	AZ468166	AZ468166 1M0281104	c	526	27	0.5	442	150	BF551292	BF551292 UI-R-CO-h
454	346	3	AA219225	AA219225 zq16g12.r	527	27	0.5	444	17	AI186252	AI186252 qd34d02.x	
455	346	119	AW707038	AW707038 sk20g08.y	c	528	27	0.5	444	136	BE519343	BE519343 945023C08
456	346	130	BB274694	BB274694 BB274694	c	529	27	0.5	445	149	BF469402	BF469402 UI-M-BH3-
457	348	138	BE665360	BE665360 154308.MA	c	530	27	0.5	446	118	AW632941	AW632941 b102a06.x
458	350	10	AA655964	AA655964 vs45a02.r	531	27	0.5	448	247	AZ639153	AZ639153 1M0499N21	
459	350	20	AI462298	AI462298 ub62a10.x	532	27	0.5	451	16	AI138555	AI138555 qd80h08.x	
460	350	163	BE120293	BE120293 UI-R-CA0-	c	533	27	0.5	451	22	AI586319	AI586319 vs45a02.x
461	352	4	AA280412	AA280412 zt05d02.s	c	534	27	0.5	451	103	AI912631	AI912631 we11h06.x
462	353	238	AZ116802	AZ116802 RPCI-23-4	535	27	0.5	451	249	AZ753044	AZ753044 RPCI-24-8	
463	355	236	AZ004375	AZ004375 RPCI-23-3	536	27	0.5	453	9	AA618552	AA618552 np38d01.s	
464	357	32	AV678291	AV678291 AV678291	537	27	0.5	455	13	AA938656	AA938656 oc08g01.s	
465	362	21	AI556038	AI556038 UI-R-C2p-	538	27	0.5	455	145	BF199384	BF199384 249531.MA	
466	363	3	AA209596	AA209596 mw75d08.r	c	539	27	0.5	456	137	BE575511	BE575511 dc57d01.x
467	365	4	AA279444	AA279444 z84g02.s	c	540	27	0.5	461	240	AZ246274	AZ246274 RPCI-23-3
468	366	240	AZ267412	AZ267412 RPCI-23-4	c	541	27	0.5	462	19	AI348771	AI348771 ta88a11.x
469	367	112	AW145699	AW145699 ga2h08.y	542	27	0.5	462	118	AW633057	AW633057 b103c10.x	
470	370	119	AW710174	AW710174 el909ne.f	c	543	27	0.5	463	156	C22642	C22642 C22642 Rice
471	371	143	BF059743	BF059743 7k65g04.x	c	544	27	0.5	463	187	N94183	N94183 za26d12.r1
472	373	118	AW589566	AW589566 mx12b08.x	545	27	0.5	465	23	AI650622	AI650622 wb01g08.x	
473	376	11	AA721019	AA721019 nx89f10.s	546	27	0.5	468	250	AZ832378	AZ832378 2M0112E15	
474	376	11	AA771567	AA771567 vm46b11.r	547	27	0.5	469	118	AW632912	AW632912 b101f12.x	
475	376	144	BF131961	BF131961 601820940	c	548	27	0.5	470	32	AV715898	AV715898 AV715898
476	381	226	AQ283388	AQ283388 RPCI111-89	c	549	27	0.5	475	146	BF286258	BF286258 EST450849
477	383	15	AI059875	AI059875 UI-R-C1-1	c	550	27	0.5	476	237	AZ068159	AZ068159 RPCI-23-4
478	385	119	AW642708	AW642708 cm21f09.w	c	551	27	0.5	477	20	AI477812	AI477812 fb55d04.y
479	386	249	AZ753512	AZ753512 RPCI-24-1	c	552	27	0.5	478	110	AV754037	AV754037 AV754037
480	388	3	AA209579	AA209579 mw75g07.r	c	553	27	0.5	478	223	AQ040554	AQ040554 CIR-HSP-2
481	388	111	AW085780	AW085780 xc55g03.x	554	27	0.5	479	237	AZ051332	AZ051332 slt00039	
482	389	32	AV723324	AV723324 AV723324	555	27	0.5	481	143	BF059431	BF059431 7K59e07.x	
483	390	10	AA678205	AA678205 z115c11.s	556	27	0.5	481	245	AZ517260	AZ517260 RPCI-11-8	
484	391	11	AA766766	AA766766 oa35c10.s	c	557	27	0.5	483	189	W03733	W03733 za74e07.r1
485	395	3	AA213318	AA213318 mw83e06.r	c	558	27	0.5	484	16	AJ126716	AJ126716 qb94d01.x
486	395	4	AA292006	AA292006 z128b11.r	c	559	27	0.5	490	102	AI806046	AI806046 te48b01.x
487	397	9	AA601154	AA601154 no54f11.s	560	27	0.5	490	104	AI968136	AI968136 wu13e10.x	
488	397	18	AI267656	AI267656 aq92g10.x	561	27	0.5	491	256	B23561	B23561 FI5G15TF IG	
489	401	22	AI604666	AI604666 vm46b11.y	c	562	27	0.5	492	116	AA435086	AA435086 UI-R-BJ0p
490	402	7	AA454486	AA454486 zw29e02.r	c	563	27	0.5	496	22	AI631896	AI631896 wa37d10.x
491	402	20	AI467453	AI467453 vd53g10.x	c	564	27	0.5	497	138	BE623310	BE623310 ut98h10.x
492	402	146	BE233441	BE233441 602024228	565	27	0.5	500	107	AU087396	AU087396 AU087396	
493	402	162	BE039664	BE039664 OC02B10.O	566	27	0.5	500	224	AQ135902	AQ135902 HS_3060_A	
494	403	158	H73304	H73304 yu27c10.r1	567	27	0.5	501	227	AQ373119	AQ373119 RPCI11-14	
495	403	187	RS6140	RS6140 yu27c10.r1	568	27	0.5	501	249	AZ768417	AZ768417 1M0568H22	
496	403	231	AQ0636476	AQ0636476 RPCI-11-4	569	27	0.5	502	243	AZ422665	AZ422665 1M0201M20	
497	405	32	AV733917	AV733917 AV733917	570	27	0.5	505	118	AW632884	AW632884 b101d08.x	
498	406	9	AA620163	AA620163 vo63c08.r	c	571	27	0.5	507	16	AI155607	AI155607 ue02d03.r
499	406	224	AQ144121	AQ144121 HS_3074_A	c	572	27	0.5	507	229	AQ459895	AQ459895 HS_5131_B
500	409	8	AA474852	AA474852 vd53g10.s	c	573	27	0.5	508	248	AZ744632	AZ744632 RPCI-24-1
501	409	138	BE625776	BE625776 ut98h10.y	c	574	27	0.5	508	251	AZ871469	AZ871469 2M018A23
502	409	168	BF687773	BF687773 602066822	575	27	0.5	509	251	AZ871469	AZ871469 2M018A23	
503	411	7	AA454820	AA454820 zx79b07.s	576	27	0.5	513	24	AI742081	AI742081 wq38g03.x	
504	411	21	AI493846	AI493846 qz51f04.x	c	577	27	0.5	513	106	AT005187	AT005187 AT005187
505	413	7	AA454809	AA454809 zx79a04.s	c	578	27	0.5	514	19	AJ399632	AJ399632 NCSC5B11T



287	28	0.5	455	23	AI689088	AI689088 tx81f08.x	C 360	27	0.5	182	110	AV742759	AV742759
288	28	0.5	458	22	AI630179	AI630179 ae26a08.r	C 361	27	0.5	191	8	AA501754	AA501754 ng05a02.s
289	28	0.5	464	16	AI132403	AI132403 uo24c05.x	C 362	27	0.5	191	18	AI251979	AI251979 qv57g09.x
290	28	0.5	476	22	AI601747	AI601747 fc04c01.x	C 363	27	0.5	191	257	T02709	T02709 0064M3 gmbp
291	28	0.5	476	107	AU096057	AU096057 AU096057	C 363	27	0.5	191	257	T02709	AI305591 qw47b08.x
292	28	0.5	477	19	AI342754	AI342754 qo26c02.x	C 364	27	0.5	192	18	AI305591	AI305591 qv47b08.x
293	28	0.5	477	188	R99446	R99446 yq79g03.r1	C 365	27	0.5	192	18	AI308401	AI308401 tb43e01.x
294	28	0.5	477	238	A2090343	A2090343 RPCI-23-4	C 366	27	0.5	193	249	A2767740	A2767740 IM0567N09
295	28	0.5	483	248	A2729776	A2729776 RPCI-24-9	C 367	27	0.5	196	4	AA250310	AA250310 mw60a12.r
296	28	0.5	486	105	AI370886	AI370886 mTB4A0806	C 368	27	0.5	200	3	AA208868	AA208868 mw75g02.r
297	28	0.5	515	175	BG249676	BG249676 602319788	C 369	27	0.5	204	157	D79598	D79598 HUM288B04B
298	28	0.5	528	244	A2492998	A2492998 IM0327119	C 370	27	0.5	207	7	AA405488	AA405488 zw39e04.s
299	28	0.5	529	105	AL501036	AL501036 AL501036	C 371	27	0.5	209	20	AI422359	AI422359 tf65h10.x
300	28	0.5	537	156	C83907	C83907 C83907 Dict	C 372	27	0.5	210	161	BB591008	BB591008 BB591008
301	28	0.5	546	240	A2259147	A2259147 RPCI-23-1	C 373	27	0.5	214	110	AW021716	AW021716 df27d01.y
302	28	0.5	547	118	AW635587	AW635587 bl35f02.w	C 374	27	0.5	215	163	BE076467	BE076467 CM0-BT059
303	28	0.5	564	231	AQ634217	AQ634217 RPCI-11-4	C 375	27	0.5	221	13	AA912444	AA912444 om52a02.s
304	28	0.5	564	244	A2456129	A2456129 IM0258K17	C 376	27	0.5	221	17	AI223756	AI223756 qx32d05.x
305	28	0.5	569	19	AI392580	AI392580 NCSMLF7T3	C 377	27	0.5	225	161	BB590643	BB590643 BB590643
306	28	0.5	575	256	B77506	B77506 T26H20TR TA	C 378	27	0.5	229	21	AI538430	AI538430 t06b05.x
307	28	0.5	584	107	AU029437	AU029437 AU029437	C 379	27	0.5	235	246	A2618385	A2618385 IM0450001
308	28	0.5	606	244	A2448954	A2448954 IM0246F22	C 380	27	0.5	236	27	AV239760	AV239760 AV239760
309	28	0.5	607	11	AA769588	AA769588 n242b06.s	C 381	27	0.5	237	106	AL5888277	AL5888277 AL5888277
310	28	0.5	612	249	AZ770203	AZ770203 IM0571E10	C 382	27	0.5	238	4	AA238681	AA238681 mx82h09.r
311	28	0.5	613	123	AW982845	AW982845 HVSMEG000	C 383	27	0.5	238	18	AI281869	AI281869 qf68g04.x
312	28	0.5	626	241	A2280050	A2280050 RPCI-23-1	C 384	27	0.5	242	127	BB151720	BB151720 BB151720
313	28	0.5	627	247	A2655154	A2655154 IM0530C02	C 385	27	0.5	242	161	BB590018	BB590018 BB590018
314	28	0.5	659	153	BG435803	BG435803 602508422	C 386	27	0.5	246	115	AW396453	AW396453 sb03h07.y
315	28	0.5	681	243	A2436700	A2436700 IM0224K05	C 387	27	0.5	247	17	AI203372	AI203372 qf30a05.x
316	28	0.5	695	243	A2420291	A2420291 IM0198M06	C 388	27	0.5	250	125	BB076877	BB076877 BB076877
317	28	0.5	745	168	BF180502	BF180502 60125294	C 389	27	0.5	251	123	AW996630	AW996630 MRO-BN007
318	28	0.5	749	168	BF699770	BF699770 602127207	C 390	27	0.5	252	4	AA245675	AA245675 mw05g11.r
319	28	0.5	782	153	BG434777	BG434777 602507988	C 391	27	0.5	252	148	BF448500	BF448500 7n71e10.x
320	28	0.5	862	172	BG025091	BG025091 602276440	C 392	27	0.5	253	27	AV260966	AV260966 AV260966
321	28	0.5	877	168	BF675571	BF675571 602083459	C 393	27	0.5	253	138	BE672323	BE672323 7a59d01.x
322	28	0.5	897	168	BF673321	BF673321 602136175	C 394	27	0.5	260	161	BB579507	BB579507 BB579507
323	28	0.5	967	150	BF577032	BF577032 602134491	C 395	27	0.5	261	189	TG0887	TG0887 yb72e04.r1
324	28	0.5	992	219	CNS0026W	CNS0026W Drosophil	C 396	27	0.5	264	19	AI376483	AI376483 tb76f06.x
325	28	0.5	1046	169	BF790409	BF790409 602249661	C 397	27	0.5	266	111	AW069860	AW069860 cr52e10.x
326	28	0.5	1129	220	CNS02ATF	AL188988 Tetraodon	C 398	27	0.5	266	18	AI311987	AI311987 ta89f11.x
327	28	0.5	1224	142	BE962061	BE962061 601655121	C 399	27	0.5	266	22	AI611438	AI611438 tt61d03.x
328	28	0.5	1287	172	BF980431	BF980431 602304159	C 400	27	0.5	267	18	AI313286	AI313286 tt62d04.x
329	28	0.5	50	118	AW632886	AW632886 bl01d10.x	C 401	27	0.5	275	19	AI335410	AI335410 tb66b01.x
330	27	0.5	54	244	A2482068	A2482068 IM0306M23	C 402	27	0.5	277	19	AI345286	AI345286 tb68b11.x
331	27	0.5	67	9	AA587263	AA587263 nn70f08.s	C 403	27	0.5	279	18	AI312292	AI312292 ta89a12.x
332	27	0.5	73	146	BP213384	BP213384 601845128	C 404	27	0.5	279	19	AI345434	AI345434 tb69e04.x
333	27	0.5	89	18	AI306886	AI306886 qw75f12.x	C 405	27	0.5	279	19	AI348772	AI348772 ta85a12.x
334	27	0.5	97	159	NS1554	NS1554 yf86a02.s1	C 406	27	0.5	283	23	AI636720	AI636720 ts57f05.x
335	27	0.5	104	4	AA270996	AA270996 va85c07.r	C 407	27	0.5	283	124	BB067718	BB067718 BB067718
336	27	0.5	105	249	A2797219	A2797219 2M0053015	C 408	27	0.5	284	2	AA125017	AA125017 mq73d07.r
337	27	0.5	116	22	AI580620	AI580620 tb72a11.x	C 409	27	0.5	284	131	BB340235	BB340235 BB340235
338	27	0.5	118	18	AI275346	AI275346 qw66d05.x	C 410	27	0.5	285	3	AA209063	AA209063 mw75h12.r
339	27	0.5	121	9	AA624364	AA624364 vn02f02.r	C 411	27	0.5	286	9	AA551610	AA551610 nj37b05.s
340	27	0.5	124	18	AI308444	AI308444 tb43h10.x	C 412	27	0.5	286	161	BB580952	BB580952 BB580952
341	27	0.5	132	18	AI308274	AI308274 tb42c04.x	C 413	27	0.5	287	22	AI613570	AI613570 tt80d08.x
342	27	0.5	135	18	AI305997	AI305997 qw74a07.x	C 414	27	0.5	288	27	AV259060	AV259060 AV259060
343	27	0.5	132	8	AA503588	AA503588 ng07f10.s	C 415	27	0.5	288	19	AI371951	AI371951 tb76b10.x
344	27	0.5	134	20	AI453746	AI453746 tj39g09.x	C 416	27	0.5	291	156	C22126	C22126 C22126 Miya
345	27	0.5	137	103	AI910910	AI910910 tq39f11.x	C 417	27	0.5	291	131	BB346822	BB346822 BB346822
346	27	0.5	139	122	AW885758	AW885758 RC4-OT007	C 418	27	0.5	293	12	AA853343	AA853343 NHTBcae05
347	27	0.5	143	18	AI275333	AI275333 qw66c03.x	C 419	27	0.5	298	187	R60340	R60340 yho4f03.s1
348	27	0.5	143	18	AI305997	AI305997 qw74a07.x	C 420	27	0.5	299	15	AI021015	AI021015 ua98f12.r
349	27	0.5	143	18	AI306026	AI306026 qw74d03.x	C 421	27	0.5	299	127	BB161124	BB161124 BB161124
350	27	0.5	144	18	AI312962	AI312962 ta80b01.x	C 422	27	0.5	300	149	BF459325	BF459325 UI-M-B21-
351	27	0.5	147	137	BE553062	BE553062 946088H04	C 423	27	0.5	300	230	AO582335	AO582335 RPCI-11-4
352	27	0.5	149	18	AI252259	AI252259 qv24c04.x	C 424	27	0.5	301	114	AA853343	AA853343 xrt81d12.x
353	27	0.5	152	112	AW164498	AW164498 se73e01.y	C 425	27	0.5	302	122	AW960326	AW960326 Exst372397
354	27	0.5	153	131	BB350763	BB350763 BB350763	C 426	27	0.5	302	161	BB585511	BB585511 BB585511
355	27	0.5	156	19	AI350028	AI350028 tb63g01.x	C 427	27	0.5	308	137	BE588294	BE588294 196942 BA
356	27	0.5	162	171	BF965460	BF965460 602276904	C 428	27	0.5	309	4	AA271459	AA271459 va83d07.r
357	27	0.5	166	170	BF886776	BF886776 FM4-TN017	C 429	27	0.5	309	148	BF440552	BF440552 BS2900021
358	27	0.5	169	118	AW633013	AW633013 bl102h01.x	C 430	27	0.5	309	160	BB547267	BB547267 BS247267
359	27	0.5	171	17	AI246069	AI246069 qv52f03.x	C 431	27	0.5	312	143	BF016061	BF016061 uy30c05.y
360	27	0.5	178	28	AV313767	AV313767 AV313767	C 432	27	0.5	314	23	AI651938	AI651938 mw51c12.x



C 141	30	0.6	419	11	AA710888	vt54d08.r	C 214	28	0.5	226	123	BB015139
C 142	30	0.6	447	248	AZ712606	RPCI-24-1	C 215	28	0.5	228	123	BB017258
C 143	30	0.6	496	247	AZ663714	1M0543C05	C 216	28	0.5	228	123	BB017332
C 144	30	0.6	513	107	AU029521	AU029521	C 217	28	0.5	231	27	AV254447
C 145	30	0.6	549	251	AZ889472	RPCI-24-1	C 218	28	0.5	236	27	AV269256
C 146	30	0.6	570	102	A1794456	f44f03.y	C 219	28	0.5	236	123	BB018039
C 147	30	0.6	580	23	A1666974	fc24ai12.y	C 220	28	0.5	237	28	AV281693
C 148	30	0.6	700	105	AL501171	AL501171	C 221	28	0.5	238	27	AV272794
C 149	30	0.6	830	235	AQ895992	HS_5439_A	C 222	28	0.5	238	27	AV272843
C 150	29	0.6	59	9	AA590010	vm22e01.r	C 223	28	0.5	239	30	AV371724
C 151	29	0.6	186	175	BG256231	602369194	C 224	28	0.5	241	123	BB012669
C 152	29	0.6	234	8	AA491917	ng55e06.s	C 225	28	0.5	245	123	BB016324
C 153	29	0.6	250	123	BB013584	BB013584	C 226	28	0.5	246	27	AV272589
C 154	29	0.6	306	156	C90988	C90988 Dict	C 227	28	0.5	246	162	BE020968
C 155	29	0.6	304	148	BF425689	ss03d04.y	C 228	28	0.5	248	27	AV275666
C 156	29	0.6	309	152	BG319551	1dh16 Hum	C 229	28	0.5	249	15	AI056494
C 157	29	0.6	327	164	BE155002	PM4-HT034	C 230	28	0.5	249	223	AQ029856
C 158	29	0.6	329	20	A1450427	mu61d04.y	C 231	28	0.5	250	27	AV256785
C 159	29	0.6	338	130	BB284170	BB284170	C 232	28	0.5	250	30	AV376677
C 160	29	0.6	360	247	AZ654453	1M0528F09	C 233	28	0.5	251	4	AA245902
C 161	29	0.6	373	142	BE943281	EST422860	C 234	28	0.5	251	121	AW880042
C 162	29	0.6	379	165	BE272892	601171162	C 235	28	0.5	255	28	AV280970
C 163	29	0.6	382	107	AU078389	AU078389	C 236	28	0.5	256	27	AV268782
C 164	29	0.6	395	246	AZ619458	1M0451P07	C 237	28	0.5	257	27	AV263958
C 165	29	0.6	402	141	BE895956	601438608	C 238	28	0.5	263	28	AV283264
C 166	29	0.6	418	4	AA275048	vd06h11.r	C 239	28	0.5	264	22	AI611889
C 167	29	0.6	439	228	AQ408538	HS_5132_A	C 240	28	0.5	264	27	AV265883
C 168	29	0.6	441	138	BE626435	uu05h01.y	C 241	28	0.5	266	27	AV263859
C 169	29	0.6	469	243	AZ399670	1M0165D12	C 242	28	0.5	266	161	BB576414
C 170	29	0.6	502	107	AU096183	AU096183	C 243	28	0.5	267	27	AV253518
C 171	29	0.6	505	13	AA878563	oj37b01.s	C 244	28	0.5	267	27	AV267252
C 172	29	0.6	535	120	AW792940	CW0-UM000	C 245	28	0.5	269	27	AV253547
C 173	29	0.6	550	104	AI950092	wQ05C10.x	C 246	28	0.5	269	109	AV060355
C 174	29	0.6	553	142	BE968735	601650194	C 247	28	0.5	273	27	AV253567
C 175	29	0.6	602	24	A1722145	fd19g08.y	C 248	28	0.5	275	113	AW235032
C 176	29	0.6	622	243	AZ420427	1M0198G15	C 249	28	0.5	277	17	AV259770
C 177	29	0.6	630	242	AZ374381	1M0127M03	C 250	28	0.5	279	19	AI343762
C 178	29	0.6	701	107	AU096155	AU096155	C 251	28	0.5	279	27	AV260577
C 179	29	0.6	724	256	B07038	CSRL-90D7-u	C 252	28	0.5	289	28	AV278325
C 180	29	0.6	773	233	AQ477996	HS_5537_A	C 253	28	0.5	300	27	AV258324
C 181	29	0.6	848	146	BF255756	HVSMF000	C 254	28	0.5	301	28	AV311522
C 182	29	0.6	861	248	AZ683164	ENTKP62TR	C 255	28	0.5	304	136	BE515774
C 183	29	0.6	923	221	CNS04055	AL268466	C 256	28	0.5	306	27	AV258163
C 184	29	0.6	942	220	CNS031K7	AL223648	C 257	28	0.5	309	163	BE068011
C 185	29	0.6	1461	80	BF281026	GA_EB004	C 258	28	0.5	313	27	AV258206
C 186	28	0.5	81	18	AI309402	tb37c07.x	C 259	28	0.5	319	109	AV103136
C 187	28	0.5	125	18	AI305631	qw72f07.x	C 260	28	0.5	320	125	BB099315
C 188	28	0.5	128	107	AU094731	AU094731	C 261	28	0.5	320	27	AV258016
C 189	28	0.5	142	30	AV370903	AV370903	C 262	28	0.5	327	242	AZ380474
C 190	28	0.5	156	20	AI432824	th50a03.x	C 263	28	0.5	329	18	AI312422
C 191	28	0.5	157	19	AI337758	qw87b05.x	C 264	28	0.5	336	27	AV259601
C 192	28	0.5	158	17	AI246861	qx77c05.x	C 265	28	0.5	338	27	AV260650
C 193	28	0.5	158	18	AI284747	qu24e10.x	C 266	28	0.5	359	241	AZ321275
C 194	28	0.5	158	18	AI306874	qw75e10.x	C 267	28	0.5	361	4	AA255228
C 195	28	0.5	158	18	AI308464	tb44b10.x	C 268	28	0.5	366	105	AL502036
C 196	28	0.5	160	19	AI336821	qw88d02.x	C 269	28	0.5	367	156	C96793
C 197	28	0.5	161	18	AI270889	qw67e04.x	C 270	28	0.5	367	231	AQ635444
C 198	28	0.5	161	18	AI305640	qw72g04.x	C 271	28	0.5	379	102	AI798449
C 199	28	0.5	161	18	AI306907	qw75h12.x	C 272	28	0.5	386	105	AL501433
C 200	28	0.5	170	17	AI247056	qw73c12.x	C 273	28	0.5	390	4	AA277209
C 201	28	0.5	180	18	AI289681	qw41d12.x	C 274	28	0.5	392	171	BF928725
C 202	28	0.5	181	24	AI733998	qw55g05.x	C 275	28	0.5	394	118	AW632928
C 203	28	0.5	187	17	AI247030	qw73a05.x	C 276	28	0.5	405	224	AQ148155
C 204	28	0.5	187	18	AI251238	qw55g05.x	C 277	28	0.5	407	19	AI373158
C 205	28	0.5	187	102	AI792899	qw55g05.y	C 278	28	0.5	412	240	AZ235622
C 206	28	0.5	195	162	BE049171	x73c09.x	C 279	28	0.5	416	144	BF072420
C 207	28	0.5	199	144	BF077537	227467 MA	C 280	28	0.5	418	118	AW613521
C 208	28	0.5	200	187	R38375	yc94f03.s1	C 281	28	0.5	421	155	BG545369
C 209	28	0.5	203	235	AQ090958	GSSTC0980	C 282	28	0.5	425	159	N23219
C 210	28	0.5	203	26	AV227430	AV227430	C 283	28	0.5	432	75	AW109345
C 211	28	0.5	205	25	AV107488	AV107488	C 284	28	0.5	433	162	BE054703
C 212	28	0.5	224	28	AV280977	AV280977	C 285	28	0.5	437	245	AZ511014
C 213	28	0.5	224	123	BB018851	BB018851	C 286	28	0.5	446	162	BE056257

BB015139 BB015139  
BB017258 BB017258  
BB017332 BB017332  
AV254447 AV254447  
AV269256 AV269256  
BB018039 BB018039  
AV281693 AV281693  
AV272794 AV272794  
AV272843 AV272843  
AV371724 AV371724  
BB012669 BB012669  
BB016324 BB016324  
AV272589 AV272589  
BE020968 BE020968  
AV275666 AV275666  
AI056494 AI056494  
AQ029856 AQ029856  
AV256785 AV256785  
AV376677 AV376677  
AA245902 AA245902  
AW880042 AW880042  
AV280970 AV280970  
AV268782 AV268782  
AV263958 AV263958  
AV283264 AV283264  
AI611889 AI611889  
BB576414 BB576414  
AV263859 AV263859  
AV253518 AV253518  
AV267252 AV267252  
AV253547 AV253547  
AV253567 AV253567  
AV060355 AV060355  
AV253032 AV253032  
AV259770 AV259770  
AI349762 AI349762  
AV260577 AV260577  
AV278325 AV278325  
AV258324 AV258324  
AV311522 AV311522  
BE515774 BE515774  
AV258163 AV258163  
AV258016 AV258016  
AZ380474 AZ380474  
AI312422 AI312422  
AV259601 AV259601  
AV260650 AV260650  
AZ321275 AZ321275  
AA255228 AA255228  
AL502036 AL502036  
C96793 C96793  
AQ635444 AQ635444  
AI798449 AI798449  
AL501433 AL501433  
AA277209 AA277209  
BF928725 BF928725  
AW632928 AW632928  
AQ148155 AQ148155  
AI373158 AI373158  
AZ235622 AZ235622  
BF072420 BF072420  
AW613521 AW613521  
BG545369 BG545369  
N23219 N23219  
AW109345 AW109345  
BE054703 BE054703  
AZ511014 AZ511014  
BE056257 BE056257











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:34:47 ; Search time 3712.66 Seconds  
(without alignments)  
13420.564 Million cell updates/sec

Title: US-09-512-581-1  
Perfect score: 5271  
Sequence: 1 ccggagagcccgagtgag.....gtaaaaaaaaaaaaaa 5271

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*



**THIS PAGE BLANK (USPTO)**







```
Db 736 wwwwww 743
```

RESULT 15

RESOL 10  
AAF58259  
ID AAF58259 standard; DNA; 936 BP.

AC AAF58259:

DT 24-APR-2001 (first entry)

DE Oligonucleotide D2004.

Electron-transfer group; ETM; mismatch; genotyping; XX

KW gene expression: ss.

OS - Synthetic.

AA  
PN  
WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476;

XX 26-JUL-1999: 99US-0145695.

PR I7-MAR-2000; 2000US-0190259.  
XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface.

PS Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP: 6 A: 138 C: 8 G: 8 T: 776 other:

Query Match 3.18; Score 165.6; DB 22; Length 936;

Best Local Similarity 0.88; Pred. No. 1.6e-25;

<b>Matches</b>	6;	Conservative	494;	Mismatches	228;	Indels	0;	Gaps	0;
----------------	----	--------------	------	------------	------	--------	----	------	----

Qy 4544 atacattgtggtcacatgccttagccatacacacatatggtaacatttgactatggagctctgt 4603  
::: :::: : : : : :::: : : : : : : : : : : : : : :  
Db 16 ww 75

[illegible][illegible]

Oy 4724 ctgacagataccgagctgtttattggatctattgattgaaagaatttgttaggatatagat 4783

Qy 4784 cttaagcagtaatctgcagtggtttgtatttgcgaatttactgtgaaaaa 4843  
::: : ::: : : : :: :::: |::::: : : :::::

[illegible]

•  
•  
•

Qy	4844	aattgttttcaacaattggtgtcatcttcttgatgtcactattgttggagagttaaat	4903
Db	316	:     :     :     :     :     :     :     :     :	375
Qy	4904	ggctctcccttgtgtattcttaacctagtgttactcctgggcacccctaacttcaga	4963
Db	376	:     :       :     :     :     :     :     :     :	435
Qy	4964	ggfgctaaattgtctgccattcacacagaaggatgcctctgatagaggacacaccatga	5023
Db	436	:     :     :     :     :     :     :     :     :	495
Qy	5024	aatttgaaaatagtcctgaagtcttggattactttacacctcagtatgttgtcccc	5083
Db	496	:     :     :     :     :     :     :     :     :	555
Qy	5084	gaattctgtgcctttcatcgccaatgaaattttaagaagaagaatttaaagtagttttaa	5143
Db	556	:     :     :     :     :     :     :     :     :	615
Qy	5144	ttttaagagtggttatataaataatgtactgaattctttatccattttatcatccttt	5203
Db	616	:     :     :     :     :     :     :     :     :	675
Qy	5204	cagtttttattaatctactgtatcataaaaaattctgttaatttgatgagtaaaaaaaaa	5263
Db	676	:     :     :     :     :     :     :     :     :	735
Qy	5264	aaaaaaaa 5271	
Db	736	:     :     :     :     :     :     :     :     :	743

Search completed: September 25, 2001, 19:52:23  
Job time: 19461 sec



Db 76 ..... 135  
Qy 4664 ctttttttaattgtctgattctgaagtgctgtatagctttttatctgcgcttttaa 4723  
Db 136 ..... 195  
Qy 4724 ctgacagtaccgcactgtttattgagctattgattgaaagaatttgttaggtagat 4783  
Db 196 g..... 255  
Qy 4784 cttgaagcagtaactgtcagtgctgtttgtattgtatttctgcaattttactgtgaaaaa 4843  
Db 256 ..... 315  
Qy 4844 aattgttttcaacaattggctcatttctgtgattcactatttctgtgagagtaaat 4903  
Db 316 ..... 375  
Qy 4904 ggtctctcccttctgtatcttactagtgcttactctcctggcacccttaattctcaga 4963  
Db 376 ..... 435  
Qy 4964 ggtgctaattgtctgcattacaccagaagatgcctctgtagaggagacaaccatgca 5023  
Db 436 ..... 495  
Qy 5024 aattgtgaaatagtcctgaagtcttctgattcacttctacacctcagtgatttctccca 5083  
Db 496 ..... 555  
Qy 5084 gaatttctgcttctcatggaatttgaatttgaagaagaattttaaagtagttttaa 5143  
Db 556 ..... 615  
Qy 5144 ttttaagagtggtttataaaataatgactgaattcttcccatctttatccatcttt 5203  
Db 616 ..... 675  
Qy 5204 cagtttttatttaactctactgtatcaataaattctgttaatttgatgagtaaaaaa 5263  
Db 676 ..... 735  
Qy 5264 aaaaaaa 5271  
Db 736 ..... 743

RESULT 14

AAF58257  
ID AAF58257 standard; DNA; 936 BP.  
XX  
AC AAF58257;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1954.  
XX  
KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PP 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
FA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX

PI Umek RM;  
XX WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
XX a single surface -  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 3.1%; Score 165.6; DB 22; Length 936;  
Best Local Similarity 0.8%; Pred. No. 1.6e-25;  
Matches 6; Conservative 494; Mismatches 228; Indels 0; Gaps 0;  
Qy 4544 atacatttggctacatgcttttagccatcacatggttaacattgactgagcttgt 4603  
Db 16 ..... 75  
Qy 4604 gaaagtgtaattgctgctatgtagacataaagaagaacttgaataatcttttt 4663  
Db 76 ..... 135  
Qy 4664 ctttttttaattgtctgattctgaagtgctgtatagctttttatctgcgcttttaa 4723  
Db 136 ..... 195  
Qy 4724 ctgacagtaccgcactgtttattgagctattgattgaaagaatttgttaggtagat 4783  
Db 196 g..... 255  
Qy 4784 cttgaagcagtaactgtcagtgctgtttgtattgtatttctgcaattttactgtgaaaaa 4843  
Db 256 ..... 315  
Qy 4844 aattgttttcaacaattggctcatttctgtgattcacttctacacctcagtgatttctccca 4903  
Db 316 ..... 375  
Qy 4904 ggtctctcccttctgtatcttactagtgcttactctcctggcacccttaattctcaga 4963  
Db 376 ..... 435  
Qy 4964 ggtgctaattgtctgcattacaccagaagatgcctctgtagaggagacaaccatgca 5023  
Db 436 ..... 495  
Qy 5024 aattgtgaaatagtcctgaagtcttctgattcacttctacacctcagtgatttctccca 5083  
Db 496 ..... 555  
Qy 5084 gaatttctgcttctcatggaatttgaatttgaagaagaattttaaagtagttttaa 5143  
Db 556 ..... 615  
Qy 5144 ttttaagagtggtttataaaataatgactgaattcttcccatctttatccatcttt 5203  
Db 616 ..... 675  
Qy 5204 cagtttttatttaactctactgtatcaataaattctgttaatttgatgagtaaaaaa 5263  
Db 676 ..... 735  
Qy 5264 aaaaaaa 5271



[illegible]

Qy	5084	gaatttttgccttcacatggcacaataattttaagaagaaagtatttaaatgatttttaa	514
Db	556	www.....  : : : .....  : : : .....	615
Qy	5144	ttttaagagtggtataaaaactactgaattctttcccatTTTatcaccTTt	5203
Db	616	www.....  : : : .....  : : : .....	675
Qy	5204	cagtttttatactactgtatcaataaaaaattctgtatTTGaattggaatgagTaataaaaaa	5263
Db	676	www.....  : : : .....  : : : .....	735
Qy	5264	aaaaaaaa 5271	
Db	736	www..... 743	
 RESULT 13 AAF58254			
ID	AAF58254 standard; DNA; 936 BP.		
XX AC	AAF58254;		
XX DT	24-APR-2001 (first entry)		
XX DE	Oligonucleotide D1875.		
XX KW	Electron-transfer group; ETM; mismatch; genotyping;		
XX KM	gene expression; ss.		
XX OS	Synthetic.		
XX PN	WO200107665-A2.		
XX PD	01-FEB-2001.		
XX PF	26-JUL-2000; 2000WO-US20476.		
XX PR	26-JUL-1999; 99US-0145695.		
XX PR	17-MAR-2000; 2000US-0190259.		
XX PA	(CLIN-) CLINICAL MICRO SENSORS INC.		
XX PI	Umek RM;		
XX DR	WPI; 2001-159728/16.		
XX PT	Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -		
XX PS	Example 6; Page 127; 159pp; English.		
CC CI	The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping,		
CC CC	monitoring gene expression.		
SQ	Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;		
Query Match	3.1%; Score 165.6; DB 22; Length 936;		
Best Local Similarity	0.8%; Pred. No. 1.6e-25;		
Matches	6; Conservative 494; Mismatches 228; Indels 0; Gaps		
Qy	4544	atacatttgtgcacatgcttagccacacacatgctaacaattgacctaggagtcttgt	4603
Db	16	www.....  : : : .....  : : : .....	75
Qy	4604	gaaagtgaatgtcgatggctatgtagacataaagaagaacttgtaatatcttttt	4663







```

RESULT 9
AAZ14459
ID AAZ14459 standard; cDNA; 284 BP.
XX
AC AAZ14459;
XX
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:1928.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
XX
XX WO938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080566.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1096; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ1779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 284 BP; 89 A; 43 C; 46 G; 103 T; 3 other;
XX
XX Query Match 4.9%; Score 258.8; DB 20; Length 284;
XX Best Local Similarity 97.8%; Pred. No. 3.3e-45;
XX Matches 272; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

```

---

```

QY 4970 aaattgtctgccattacaccaggaaggatgctctgtatagaggagacaccatgcaaatgt 5029
Db 1 aaattgtctgccattacaccaggaaggatgctctgtatagaggagacaccatgcaaatgt 60
QY 5030 gaaatagctctgaagtctctggattactttacacctcagttatgtgtccccaattt 5089
Db 61 gaaatagctctgaagtctctggattactttacacctcagttatgtgtccccaattt 120
QY 5090 tctggcccttcattggcaatgaaaaattttaagaagaagaagatttaaaagtattttaatttaa 5149
Db 121 tctggcccttcattggcaatgaaaaattttaagaagaagaagatttaaaagtattttaatttaa 180
QY 5150 agagtgtgttataaaaataatgtactgaattctttatcccatctttatcatccttccagttt 5209
Db 181 agagtgtgttataaaaataatgtactgaattctttatcccatctttatcatccttccagttt 240
QY 5210 ttataatctactgtatcaataaaaattctgtaatttga 5247
Db 241 ttataatctactgtatca--taaatctgtaantnga 276

```

---

```

RESULT 10
AAF18338
ID AAF18338 standard; DNA; 1217 BP.
XX
AC AAF18338;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polynucleotide sequence SEQ ID 357.
XX
XX Human; lung cancer associated protein; neuroprotective; cytoskeletal;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
XX P-PSDB; AAB58462.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer.
XX
XX Claim 1; Page 815; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX associated proteins and polynucleotide sequences, their agonists, and
XX antagonists may have neuroprotective; cytoskeletal; cardioactive;
XX immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX general; nephrotropic; antiinfective; gynecological; or antibacterial
XX activity. The invention also includes antibodies specific for the
XX protein or polynucleotide sequences. The lung cancer associated
XX polynucleotide sequences may be used for detection of lung cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The proteins may be used to treat
XX disorders such as neural, immune, muscular, reproductive,

```



PR 28-APR-1998; 98US-0069047.  
PA (GEST ) GENSET.  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-038446/03.  
DR P-PSDB; AAY65247.  
XX Novel secreted protein 5' expressed sequence tag sequences used in  
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures  
XX  
PS Claim 1; Page 495; 837pp; English.  
XX  
CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)  
CC sequences, corresponding to human secreted proteins. AAY64651 to  
CC AAY65438 represent the EST-related proteins corresponding to AA242265 to  
CC AA243052. The 5' ESTs can be used for producing secreted human gene  
CC products. They can be used to identify and isolate 5' untranslated  
CC regions (UTRs) and upstream regulatory regions which control the  
CC location, development stage, rate, and quantity of protein synthesis, as  
CC well as stability of mRNA. The ESTs are also useful as probes for  
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can  
CC also be used in forensic procedures to identify individuals, or in  
CC diagnostic procedures to identify individuals having genetic diseases  
CC resulting from abnormal gene expression. The products may also be used in  
CC gene therapy protocols. The nucleic acids encoding signal peptides can be  
CC used for directing extracellular secretion of a polypeptide or the  
CC insertion of a polypeptide into a membrane, or importing a polypeptide  
CC into a cell. The proteins encoded by the EST sequences may be useful in  
CC treating a variety of human conditions. Secreted proteins have  
CC therapeutic value, and the identification of new secreted proteins is  
CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 351 BP; 107 A; 63 C; 80 G; 99 T; 2 other;

Query Match 6.3%; Score 333; DB 21; Length 351;  
Best Local Similarity 98.6%; Pred. No. 1e-60; Indels 1; Gaps 1;  
Matches 344; Conservative 2; Mismatches 2;  
QY 19 agcggagtagcagtcggaacccggagggtagaataatttctgtcatggtcattcaa 78  
DB 1 agcggagtagcagtcggaacccggagggtagaataatttctgtcatggtcattcaa 59  
QY 79 agactaggaccatgatggaaataatacatatccgcctggggtcaagaaatcatagata 138  
DB 60 agactaggaccatgatggaaataatacatatccgcctggggtcaagaaatcatagata 119  
QY 139 aaatatctaaaggagatggtgagacgattaaagatggttgaaactttttatggata 198  
DB 120 aaatatctaaaggagatggtgagacgattaaagatggttgaaactttttatggata 179  
QY 199 tggaccagactctgaagaagaaaggagctttatttaaacctagctttacattctgctt 258  
DB 180 tggaccagactctgaagaagaaaggagctttatttaaacctagctttacattctgctt 239  
QY 259 cagatttttttcaagcatcctggttaagatggttcgcttactgtagctgctgcttg 318  
DB 240 cagatttttttcaagcatcctggttaagatggttcgcttactgtagctgctgcttg 299  
QY 319 ctgatttttcaggatttatgctctgaagctccctacacatccctga 367  
DB 300 ctgatttttcaggatttatgctctgaagctccctacacatccctga 348

RESULT 8  
ID AAC02817  
XX AAC02817 standard; cDNA; 295 BP.  
AC AAC02817;  
XX

DT 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 2815.  
DE  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
OS  
XX EPI033401-A2.  
PW  
XX 06-SEP-2000.  
PD  
XX 21-FEB-2000; 2000EP-0200610.  
PF  
XX 26-FEB-1999; 99US-0122487.  
PR  
XX (GEST ) GENSET.  
PA  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI  
XX WPI; 2000-500381/45.  
DR P-PSDB; AAG02811.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SQ Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

Query Match 5.5%; Score 292; DB 21; Length 295;  
Best Local Similarity 100.0%; Pred. No. 3.7e-52;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2562 gtctgatggctacttggaatgaaataataatcacagtaaatcacaggaggaataatga 2621  
DB 1 gtctgatggctacttggaatgaaataataatcacagtaaatcacaggaggaataatga 60  
QY 2622 ttgtctaaacaataattgcattgtagtgagacttgacagaaacaggaggaataatga 2681  
DB 61 ttgtctaaacaataattgcattgtagtgagacttgacagaaacaggaggaataatga 120  
QY 2682 ccagatatctacgctgagacttgctgctgggagtgctattgtgagctggcacaagaa 2741  
DB 121 ccagatatctacgctgagacttgctgctgggagtgctattgtgagctggcacaagaa 180  
QY 2742 cccgtgtaccatgaaatcatcattagaaataatcagctatgtgcatagctatcaac 2801  
DB 181 cccgtgtaccatgaaatcatcattagaaataatcagctatgtgcatagctatcaac 240  
QY 2802 gatgaatgctatcaagtaagacaagtgttgcgcagaaacttcacaagggc 2853  
DB 241 gatgaatgctatcaagtaagacaagtgttgcgcagaaacttcacaagggc 292







XX PD XX PF XX PR PR PR PR PR XX PA PA XX PI PI PI

Escobedo J, Garcia FD, Garcia V, Grese N, Jimis MA;  
Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
Stache-Grain B, Sudduth-Klinger J, Williams LT;

PT Stacine-Crain B, Sudouth-Klinger J, Williams Li;  
 FX  
 DR WPI; 1999-494092/41.  
 XX  
 PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types

PS Claim 1: Page 1324: 2479pp: English.  
 PS differentially expressed in different  
 PT novel human genes and their express  
 F1 differentially expressed in different

The present invention describes a library of human polynucleotides comprising the sequences given in AA212532 to AA217779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA212532 to AA217779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAG12532 to AAG17779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as; cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA21232 to AA21779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

of the 5248 polynucleotide sequences given in AA212532 to AA217779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

be used for a variety of purposes, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as;  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

identifying a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antaonists.

CC diagnosis, prognosis and management of colorectal cancer/ breast cancer/  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;  
XX  
SQ

Query Match 7.8%; Score 409.4; DB 20; Length 738;  
Best Local Similarity 83.0%; Pred. No. 1.3e-76;  
Matches 563; Conservative 0; Mismatches 77; Indels 38;

Matches	Conservative	Mismatches	Indels	Gaps
563	0	77	38	7

Qy 3890 agaaaatgaagatgaacagatagtcgcgaataaagggtaaaaagagggccagaccacccaaa 3949

Db 51 agaaaatgaagatgaacagatagtcgcgaataaagggtaaaaagagggccagaccacccaaa 110

51 aaaaaatgaagatgaacagaaatagtcctcaaaaaaaatataaaagagaccccaaa

Qy	3950	acctcttggtaggttacaccataaaagagccacaatgaaaactctcaaaaaagggaag 
Db	170	acctcttggtaggttacaccataaaagagccacaatgaaaactctcaaaaaagggaag 

I/O

DB III accctcctggttgcagttacaccacaaatgaaggccaaccaattctaaaagttaa I/V

4005  
 4006  
 4007  
 4008  
 4009  
 4010  
 4011  
 4012  
 4013  
 4014  
 4015  
 4016  
 4017  
 4018  
 4019  
 4020  
 4021  
 4022  
 4023  
 4024  
 4025  
 4026  
 4027  
 4028  
 4029  
 4030  
 4031  
 4032  
 4033  
 4034  
 4035  
 4036  
 4037  
 4038  
 4039  
 4040  
 4041  
 4042  
 4043  
 4044  
 4045  
 4046  
 4047  
 4048  
 4049  
 4050  
 4051  
 4052  
 4053  
 4054  
 4055  
 4056  
 4057  
 4058  
 4059  
 4060  
 4061  
 4062  
 4063  
 4064  
 4065  
 4066  
 4067  
 4068  
 4069  
 4070  
 4071  
 4072  
 4073  
 4074  
 4075  
 4076  
 4077  
 4078  
 4079  
 4080  
 4081  
 4082  
 4083  
 4084  
 4085  
 4086  
 4087  
 4088  
 4089  
 4090  
 4091  
 4092  
 4093  
 4094  
 4095  
 4096  
 4097  
 4098  
 4099  
 4100  
 4101  
 4102  
 4103  
 4104  
 4105  
 4106  
 4107  
 4108  
 4109  
 4110  
 4111  
 4112  
 4113  
 4114  
 4115  
 4116  
 4117  
 4118  
 4119  
 4120  
 4121  
 4122  
 4123  
 4124  
 4125  
 4126  
 4127  
 4128  
 4129  
 4130  
 4131  
 4132  
 4133  
 4134  
 4135  
 4136  
 4137  
 4138  
 4139  
 4140  
 4141  
 4142  
 4143  
 4144  
 4145  
 4146  
 4147  
 4148  
 4149  
 4150  
 4151  
 4152  
 4153  
 4154  
 4155  
 4156  
 4157  
 4158  
 4159  
 4160  
 4161  
 4162  
 4163  
 4164  
 4165  
 4166  
 4167  
 4168  
 4169  
 4170  
 4171  
 4172  
 4173  
 4174  
 4175  
 4176  
 4177  
 4178  
 4179  
 4180  
 4181  
 4182  
 4183  
 4184  
 4185  
 4186  
 4187  
 4188  
 4189  
 4190  
 4191  
 4192  
 4193  
 4194  
 4195  
 4196  
 4197  
 4198  
 4199  
 4200  
 4201  
 4202  
 4203  
 4204  
 4205  
 4206  
 4207  
 4208  
 4209  
 4210  
 4211  
 4212  
 4213  
 4214  
 4215  
 4216  
 4217  
 4218  
 4219  
 4220  
 4221  
 4222  
 4223  
 4224  
 4225  
 4226  
 4227  
 4228  
 4229  
 4230  
 4231  
 4232  
 4233  
 4234  
 4235  
 4236  
 4237  
 4238  
 4239  
 4240  
 4241  
 4242  
 4243  
 4244  
 4245  
 4246  
 4247  
 4248  
 4249  
 4250  
 4251  
 4252  
 4253  
 4254  
 4255  
 4256  
 4257  
 4258  
 4259  
 4260  
 4261  
 4262  
 4263  
 4264  
 4265  
 4266  
 4267  
 4268  
 4269  
 4270  
 4271  
 4272  
 4273  
 4274  
 4275  
 4276  
 4277  
 4278  
 4279  
 4280  
 4281  
 4282  
 4283  
 4284  
 4285  
 4286  
 4287  
 4288  
 4289  
 4290  
 4291  
 4292  
 4293  
 4294  
 4295  
 4296  
 4297  
 4298  
 4299  
 4300  
 4301  
 4302  
 4303  
 4304  
 4305  
 4306  
 4307  
 4308  
 4309  
 4310  
 4311  
 4312  
 4313  
 4314  
 4315  
 4316  
 4317  
 4318  
 4319  
 4320  
 4321  
 4322  
 4323  
 4324  
 4325  
 4326  
 4327  
 4328  
 4329  
 4330  
 4331  
 4332  
 4333  
 4334  
 4335  
 4336  
 4337  
 4338  
 4339  
 4340  
 4341  
 4342  
 4343  
 4344  
 4345  
 4346  
 4347  
 4348  
 4349  
 4350  
 4351  
 4352  
 4353  
 4354  
 4355  
 4356  
 4357  
 4358  
 4359  
 4360  
 4361  
 4362  
 4363  
 4364  
 4365  
 4366  
 4367  
 4368  
 4369  
 4370  
 4371  
 4372  
 4373  
 4374  
 4375  
 4376  
 4377  
 4378  
 4379  
 4380  
 4381  
 4382  
 4383  
 4384  
 4385  
 4386  
 4387  
 4388  
 4389  
 4390  
 4391  
 4392  
 4393  
 4394  
 4395  
 4396  
 4397  
 4398  
 4399  
 4400  
 4401  
 4402  
 4403  
 4404  
 4405  
 4406  
 4407  
 4408  
 4409  
 4410  
 4411  
 4412  
 4413  
 4414  
 4415  
 4416  
 4417  
 4418  
 4419  
 4420  
 4421  
 4422  
 4423  
 4424  
 4425  
 4426  
 4427  
 4428  
 4429  
 4430  
 4431  
 4432  
 4433  
 4434  
 4435  
 4436  
 4437  
 4438  
 4439  
 4440  
 4441  
 4442  
 4443  
 4444  
 4445  
 4446  
 4447  
 4448  
 4449  
 4450  
 4451  
 4452  
 4453  
 4454  
 4455  
 4456  
 4457  
 4458  
 4459

20 I / T CCCCCCCCCCCCHHHCCHHHCCCCCCHH  
21 CCCCCCCCCCCCHHHCCHHHCCCCCCHH

22

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

-----

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525



QY	2007	gcaggctctgaaactgcttaagtgactctcaattacacatcccatctctcattcttcgtct	2066
Db	662	tcaggacttgaactctcttaagttctgtctttttcacatctcaactgcttccactctgca	721
QY	2067	gaacatttgaactcattactcggcttgctcgaaaaatggatgatgaaaaagtagcagaagct	2126
Db	722	gagacatacgactccttgctacagtccttaagaatggaggatgacaggtagcagaagct	781
QY	2127	gcactacaaattttcaaaacacaggaagcaaaattgaagaggattttccacacatcaga	2186
Db	782	gctattcaaaatttttgaataacaggtcacaaaatagaaacagacaccttcccagatacga	841
QY	2187	tcagccttgctctcgtgttttatcatcacaaatctaaaaaaggacccccctctcaagccaaa	2246
Db	842	tcgaccttaattcccattttatcatcaaaagcaagagggttactccacaccaagcaaaa	901
QY	2247	tatgccattcattgatcatcgcgtatatttctagtaaaagagacccagtttgacaagata	2306
Db	902	caggctgtgcactgtatacacgccatactcaaataaagaagtcaggctgtgcacagatt	961
QY	2307	tttgagcctctgcataagagcctagatccaagcaacctggaaactctcataaacacacttg	2366
Db	962	tttgagcctaactcagtaggagctgaatgctgatgtgccagaaacattataactccatta	1021
QY	2367	gttactattggtctaatgtctctcttgcacctgatcaatttgcgtctcttggaaatct	2426
Db	1022	gtttcattggccaatcttctatgttagcacagatcagtttgccttcccacatgaaatct	1081
QY	2427	tgggtagctacitcttcatgtgaaaatcttctcatgaatgatcggcttcaggggaaaaag	2486
Db	1082	gtatgagcaaatcttatgtgaaatctgcttaataagatgacaggtccaacaggtgaaaag	1141
QY	2487	acaaactaaactttgggttcagatgaagaagtatctctcgagacaaatgggtcaaaaattcag	2546
Db	1142	aatggaaaacttggtctccagatgaagagggttccccctgaagtaactagcaaaaggctacag	1201
QY	2547	gctattaaaaatgatggttcgaatggctacttggaatgaaaaataatcacagtaaatcagga	2606
Db	1202	gcaattaaactcttggttaagggtggctgttgggtatgaaaaaacccagctctaaatctgcc	1261
QY	2607	actctacctctaagattgctaaacaataattgcatagtgatggagacttgacagaacag	2666
Db	1262	aatcaacctctcgggttatacagcgatgttggtagtgagggtgacctgacagagcaa	1321
QY	2667	gggaaaaattagtaaacccagatatgtccagctctgagacttgcctgctggagtgctattgtg	2726
Db	1322	aagagatcagtaaatctgatatgtctcgtctggattagctgctggtagtgccataatg	1381
QY	2727	aagctggcacagaaccttgtaccatgaaatctcacatctagacaataatcagctatgt	2786
Db	1382	aagcttgctcaggaaacctgttaccatgaaattattccccagaaacagtttccagctctg	1441
QY	2787	gcattagctatacagatgaaatgctatacaatgaagacaagtggttggccagaaaacttcaac	2846
Db	1442	gcaacttgctataatgatgagtggttaccacatgaaggcagataatttgcctcagaagctgcat	1501
QY	2847	aaaggccttccgctttacggtctccacttgagtatatggcaatctgtgcccctttgtgca	2906
Db	1502	aaggcaacttgtaagttactgctccactggagtatatggcgactttgctgtgtgtgcc	1561
QY	2907	aaagaatctgtaaaaggagaaagagctcatgctaggcgaatgtttgggtgaaaaataataat	2966
Db	1562	aaagaatctgtgaaggagaaagagcacacgcagacaaatgtttacatgaaaaataatcagt	1621
QY	2967	gtaagcggaggatctatgaagcagcatgcagctgttagtgaaaaattattgtctcttcta	3026
Db	1622	atacgcagggaatacataaagcagaatcctatggctactgagaatctattatcacgttg	1681
QY	3027	ccagagtagttgttccatatacaattcaccttttggcacatgcccgagattatgtcaaa	3086
Db	1682	cctggaatatgtagttccatacatgatccactgttagccccatgatccagattttcaaga	1741
QY	3087	gtacaggatattgaacaacttaaaagatgtaaagaatgtcttggtttggcttctggaaata	3146

## RESULTS

AAV8/629  
ID AAV87629 standard; cDNA; 439 BP.

AA  
AC AAV87629;

DT 12-FEB-1999 (first entry)

EST clone DY17.

Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW  
tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW  
chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW  
receptor; ligand; anti-inflammation; tumour inhibitor; ds

OS Homo sapiens.

XX PN WQ9845437-A2

XX  
PD 15-OCT-1998.XX  
PF 10-APR-1998; 98WO-US06956.

AA 10-APR-1997; 97US-0837312.

PA (GEMY ) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

DR WPI; 1999-070078/06.

New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 126; 641pp; English.











Db	865	aggtctcgtatgcttgggaaaaacatctatcagcgaattgtcagagcatgcttttgcactaa	924
Qy	841	ttttggagctcaacaattatgatagctcatttgcctctctgttttaccocagcttgaat	900
Db	925	ttttggagctcacaattatgatagctcatttgcctctctgttttaccocagcttgaat	984
Qy	901	ttaattaaagagcaatgatataatgaggagcctacaagttgtttaaactactgccaataa	960
Db	985	ttaattaaagagcaatgatataatgaggagcctacaagttgtttaaactactgccaataa	1044
Qy	961	tgtttggggcaaggaattcagaaattggcttctcaaaaaacaccactttggcagtgctact	1020
Db	1045	tgtttggggcaaggaattcagaaattggcttctcaaaaaacaccactttggcagtgctact	1104
Qy	1021	tggcaggtttaatgataccatgtaaccaatccgctggaattggtgaaatttgcctagcc	1080
Db	1105	tggcaggtttaatgataccatgtaaccaatccgctggaattggtgaaatttgcctagcc	1164
Qy	1081	attgtctcatgaaccactcctgatttagcaaaaagacttaacagagatctcttaaagtgggt	1140
Db	1165	attgtctcatgaaccactcctgatttagcaaaaagacttaacagagatctcttaaagtgggt	1224
Qy	1141	cacatgacctggaggagctatttagacatgattgtttatgttccaatagttacagctgcta	1200
Db	1225	cacatgacctggaggagctatttagacatgattgtttatgttccaatagttacagctgcta	1284
Qy	1201	aaaaggaatatctcttggccaatgatcacctacttaatttgttgagagagagaacattag	1260
Db	1285	aaaaggaatatctcttggccaatgatcacctacttaatttgttgagagagagaacattag	1344
Qy	1261	acaaacgtaggagtagtcacaaagaagccatgatgggacttgcaccaattataagaagt	1320
Db	1345	acaaacgtaggagtagtcacaaagaagccatgatgggacttgcaccaattataagaagt	1404
Qy	1321	atgctttacagtcagcagctggaagaatgctgcacaaacagatagcatggaatcaagaaca	1380
Db	1405	atgctttacagtcagcagctggaagaatgctgcacaaacagatagcatggaatcaagaaca	1464
Qy	1381	aattgctcacatatattatccaataatgattgatgatcgactacttgttgaacggatct	1440
Db	1465	aattgctcacatatattatccaataatgattgatgatcgactacttgttgaacggatct	1524
Qy	1441	tgtctcaatacatggttctccacaatttagaaactacagaacgcatgaaatgcttatatt	1500
Db	1525	tgtctcaatacatggttctccacaatttagaaactacagaacgcatgaaatgcttatatt	1584
Qy	1501	acttgtagccacactggatttaaatgctgtgaaagcaattgaatgaaaaatgtgaaatgtgc	1560
Db	1585	acttgtagccacactggatttaaatgctgtgaaagcaattgaatgaaaaatgtgaaatgtgc	1644
Qy	1561	aaaatctgtccgacatcaagttaaaggatttgccttgacttgattaaagcaacccaacag	1620
Db	1645	aaaatctgtccgacatcaagttaaaggatttgccttgacttgattaaagcaacccaacag	1704
Qy	1621	atgccagtgccaaggccatatatttccaaaagtgatggttattacagaataatttaacctgac	1680
Db	1705	atgccagtgccaaggccatatatttccaaaagtgatggttattacagaataatttaacctgac	1764
Qy	1681	ctggtgaaggtccaggatttcaatgaagaaattccaacaggtgttgagaagatgatgagaaa	1740
Db	1765	ctggtgaaggtccaggatttcaatgaagaaattccaacaggtgttgagaagatgatgagaaa	1824
Qy	1741	taagaaagcaggttagaagctacttggtagccaaatgctcctgaagcaggctgaaggtt	1800
Db	1825	taagaaagcaggttagaagctacttggtagccaaatgctcctgaagcaggctgaaggtt	1884
Qy	1801	gtgtgcgtgaaataactaagaagttgggcaaccccaaacacagctacaaatccttctctgg	1860
Db	1885	gtgtgcgtgaaataactaagaagttgggcaaccccaaacacagctacaaatccttctctgg	1944
Qy	1861	aaatgatcaagtttctcttgggagggatagcacttgtgcacatagatcacogaatctatca	1920
Db	1945	aaatgatcaagtttctcttgggagggatagcacttgtgcacatagatcacogaatctatca	2004

QY	1921	gtgctcttattaaacaagtgaacaaatcaatagatggaacagcagatgatgaagatgagg	1981
Db	2005	gtgctcttattaaacaagtgaacaaatcaatagatggaacagcagatgatgaagatgagg	2064
QY	1981	gtgttccaaactgatcaagccatcagagcaggctctgaaactgcttaagggtactctcattta	2040
Db	2065	gtgttccaaactgatcaagccatcagagcaggctctgaaactgcttaagggtactctcattta	2124
QY	2041	caatccccatctcatttcattctgctgaaacatttgaaatcattactgctgttcttgaaaa	2100
Db	2125	caatccccatctcatttcattctgctgaaacatttgaaatcattactgctgttcttgaaaa	2184
QY	2101	tggtatgataaaaagttagcagaagctgcactcaaaattttcaaaaacacaggaagcaaaa	2160
Db	2185	tggtatgataaaaagttagcagaagctgcactcaaaattttcaaaaacacaggaagcaaaa	2244
QY	2161	tgaagagagattttccacacatcagatcagccttgctgctctgttttaccatcacaatatcta	2220
Db	2245	tgaagagagattttccacacatcagatcagccttgctgctctgttttaccatcacaatatcta	2304
QY	2221	aaaaagagccccccgtcaagccaaatatgccattcatgttaccatgcgatatttttcta	2280
Db	2305	aaaaagagccccccgtcaagccaaatatgccattcatgttaccatgcgatatttttcta	2364
QY	2281	gtaaagagaccagtttgacagatatttgagcctctgcataaagagcctagatcccaagca	2340
Db	2365	gtaaagagaccagtttgacagatatttgagcctctgcataaagagcctagatcccaagca	2424
QY	2341	acctggaaatctctataacacccattggttactatttgggtcattatgtctctcttgcacctg	2400
Db	2425	acctggaaatctctataacacccattggttactatttgggtcattatgtctctcttgcacctg	2484
QY	2401	atcaatttgtctctcttgagaactcttggtgagctacttctatgttgaaagatcttctca	2460
Db	2485	atcaatttgtctctcttgagaactcttggtgagctacttctatgttgaaagatcttctca	2544
QY	2461	tgaatgatcggtcttcaggggaaaaagacaaactaaaccttgggtttccagatgaagaagtat	2520
Db	2545	tgaatgatcggtcttcaggggaaaaagacaaactaaaccttgggtttccagatgaagaagtat	2604
QY	2521	ctcttgagacaatatggttcaaaaattcagggtctattaaaaatgatggttcgatggctacttggaa	2580
Db	2605	ctcttgagacaatatggttcaaaaattcagggtctattaaaaatgatggttcgatggctacttggaa	2664
QY	2581	tgaaaaaataatcacagtataatcaggaaacttctaccttaagattgcttaaacacaatatctgc	2640
Db	2665	tgaaaaaataatcacagtataatcaggaaacttctaccttaagattgcttaaacacaatatctgc	2724
QY	2641	atagtgtatggagacttgacagaacaggggaaaattagtaaacacagatatgtcacgtctga	2700
Db	2725	atagtgtatggagacttgacagaacaggggaaaattagtaaacacagatatgtcacgtctga	2784
QY	2701	gacttgtctgtggagtgctatttgtgaagctggcacaagaacccctgttaccatgaaatca	2760
Db	2785	gacttgtctgtggagtgctatttgtgaagctggcacaagaacccctgttaccatgaaatca	2844
QY	2761	tcacattagaacaataatcagctatgtgcattagctatcaacgatgaatgctatcaagtaa	2820
Db	2845	tcacattagaacaataatcagctatgtgcattagctatcaacgatgaatgctatcaagtaa	2904
QY	2821	gacaagtgtttgcccagaaacttcacaaagcctttcccggttacggcttccacttgagt	2880
Db	2905	gacaagtgtttgcccagaaacttcacaaagcctttcccggttacggcttccacttgagt	2964
QY	2881	atatggcaatctgtgccccttttgtcaaaaagatcctgttaaaaggagagaagagctcatgcta	2940
Db	2965	atatggcaatctgtgccccttttgtcaaaaagatcctgttaaaaggagagaagagctcatgcta	3024
QY	2941	ggcaatgtttgtgaaaaataataatgtgaaggcgggagtatcttgaagcagcatgcaagctg	3000
Db	3025	ggcaatgtttgtgaaaaataataatgtgaaggcgggagtatcttgaagcagcatgcaagctg	3084



FT exon 2625..2761  
 FT /tag= v  
 FT /number= 23  
 FT 2672..2885  
 FT /tag= w  
 FT /number= 24  
 FT 2886..3090  
 FT /tag= x  
 FT /number= 25  
 FT 3091..3205  
 FT /tag= y  
 FT /number= 26  
 FT 3206..3338  
 FT /tag= z  
 FT /number= 27  
 FT 3339..3458  
 FT /tag= aa  
 FT /number= 28  
 FT 3459..3521  
 FT /tag= ab  
 FT /number= 29  
 FT 3522..3667  
 FT /tag= ac  
 FT /number= 30  
 FT 3668..3773  
 FT /tag= ad  
 FT /number= 31  
 FT 3774..4213  
 FT /tag= ae  
 FT /number= 32  
 FT 4214..4438  
 FT /tag= af  
 FT /number= 33  
 FT 4314..4408  
 FT /tag= ag  
 FT /note= "Destabilizing signal"  
 FT 4439..5337  
 FT /tag= ah  
 FT /number= 34  
 FT 5312..5317  
 FT /tag= ai  
 FT 5333..5337  
 FT /tag= aj  
 FT /note= "Poly adenylation cleavage site"  
 XX  
 PN W0200050454-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 XX  
 PF 24-FEB-2000; 2000WO-US04732.  
 XX  
 PR 24-FEB-1999; 99US-0121461.  
 XX  
 XX  
 PA (TUFT ) TUFTS COLLEGE.  
 XX  
 PI Soto AM, Sonnenschein C, Geck P, Szelei J;  
 XX  
 DR WPI; 2000-565451/52.  
 DR P-PSDB; AAY94702.  
 XX  
 XX New human androgen-induced tumor suppressor cDNA sequence termed  
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient  
 PT diagnosis and treatment of prostate cancer -  
 XX  
 PS Example 4; Fig 6; 152pp; English.  
 XX  
 CC This invention relates to a human androgen-induced tumour suppressor  
 CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is  
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in  
 CC inhibiting cell proliferation and use as a marker for the efficient  
 CC diagnosis and treatment of prostate cancer. The invention includes AS3  
 CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host  
 CC cell transfected with the expression vector, and a method for producing

CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has  
 CC cytoskeletal activity, and acts to suppress cell proliferation. The AS3  
 CC gene is useful as a marker for the efficient diagnosis and treatment of  
 CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of  
 CC antisense agents for sequence specific modulation of gene expression. The  
 CC AS3 protein may be used in the treatment of disorders caused by aberrant  
 CC modification or mutation of a gene encoding an AS3 protein, misregulation  
 CC of the AS3 gene or aberrant post-translational modification of the AS3  
 CC protein. This sequence represents the human AS3 cDNA sequence with an  
 CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when  
 CC compared with the claimed AS3 cDNA sequence AAA28051.  
 XX  
 SQ Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 5271; DB 21; Length 5355;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 cggagagcccgagtgagcgagtagcgagtcggcaaccgaggggtagaatatatt 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 85 cggagagcccgagtgagcgagtagcgagtcggcaaccgaggggtagaatatatt 144  
 QY 61 ctgtcatggctcattcaagactagagaccatgatgaaaaattacatccgctggg 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 145 ctgtcatggctcattcaagactagagaccatgatgaaaaattacatccgctggg 204  
 QY 121 tcaaggaaatcacagataaaatctaaaggagagtggtgagacgattaaagatggtg 180  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 205 tcaaggaaatcacagataaaatctaaaggagagtggtgagacgattaaagatggtg 264  
 QY 181 tgaaaacttttatggatggaaccagactctgaagaagaaaaaggctttattaaacc 240  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 265 tgaaaacttttatggatggaaccagactctgaagaagaaaaaggctttattaaacc 324  
 QY 241 tagctttacatctgtctcagatttttttccaagcactcctggaagatgttcgcttac 300  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 325 tagctttacatctgtctcagatttttttccaagcactcctggaagatgttcgcttac 384  
 QY 301 tggtagcctgtgcttgcgtgatatatttcaggatttatgctcctgaagctccttacacat 360  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 385 tggtagcctgtgcttgcgtgatatatttcaggatttatgctcctgaagctccttacacat 444  
 QY 361 ccctgataaactaaaggatatttatgtttatacaagaacagttgaaggggtagagg 420  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 445 ccctgataaactaaaggatatttatgtttatacaagaacagttgaaggggtagagg 504  
 QY 421 atacaagagcccaaatcaatagttatttttatttacttgagaacattgcttgggtca 480  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 505 atacaagagcccaaatcaatagttatttttatttacttgagaacattgcttgggtca 564  
 QY 481 agtcataaacatagctttgagttagaagatagcaatgaaattttcaccagctataca 540  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 565 agtcataaacatagctttgagttagaagatagcaatgaaattttcaccagctataca 624  
 QY 541 gaaccttatttcagttataaaacaatggccacaatcagaagtcacatgacatggtag 600  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 625 gaaccttatttcagttataaaacaatggccacaatcagaagtcacatgacatggtag 684  
 QY 601 acctatgagctctattatttggaagtgatcacagtgctcagagcttttgatcagg 660  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 685 acctatgagctctattatttggaagtgatcacagtgctcagagcttttgatcagg 744  
 QY 661 ttttagtaaatctgtgacctgctcataagaattttaacaagcagcatatgatttggcaa 720  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 745 ttttagtaaatctgtgacctgctcataagaattttaacaagcagcatatgatttggcaa 804  
 QY 721 aggccttactgaagagagacgctcaagctattgagccatattaccacttttttaac 780  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 805 aggccttactgaagagagacgctcaagctattgagccatattaccacttttttaac 864  
 QY 781 aggttctgactgttggaagaacatctacgatttgcagagcatgtcttgcactaa 840  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||











Dbb 1 ccgagagcccccaggtgagcggagtagcagtgagtcggccaaacccggaggggtagaataattt 60  
Qy 61 ctgtcattggtcattcaaaactaggaccacatgatgaaaaattacatatccgcctggg 120  
Dbb 61 ctgtcattggtcattcaaaactaggaccacatgatgaaaaattacatatccgcctggg 120  
Qy 121 tcaaggaaatcacagataaaatattctaaagaggagatggtagacgattaaagatggtg 180  
Dbb 121 tcaaggaaatcacagataaaatattctaaagaggagatggtagacgattaaagatggtg 180  
Qy 181 tgaaaaattttatggatatggaccagactctgaagagaaaaaggagctttatttaaacc 240  
Dbb 181 tgaaaaattttatggatatggaccagactctgaagagaaaaaggagctttatttaaacc 240  
Qy 241 tagctttacatctgcttcagattttttctcaagcactcgtgaaagatgttcgcttac 300  
Dbb 241 tagctttacatctgcttcagattttttctcaagcactcgtgaaagatgttcgcttac 300  
Qy 301 tggtagcctgctgcttgcgtgatattttcaggatttatgctcctgaagctccttacacat 360  
Dbb 301 tggtagcctgctgcttgcgtgatattttcaggatttatgctcctgaagctccttacacat 360  
Qy 361 cccctgataaactaaaggatattttatgtttatacaagacagctgaaaggggctagagg 420  
Dbb 361 cccctgataaactaaaggatattttatgtttatacaagacagctgaaaggggctagagg 420  
Qy 421 atacaagagcccaacattcaataggtattttatttacttgagaacattgcttgggtca 480  
Dbb 421 atacaagagcccaacattcaataggtattttatttacttgagaacattgcttgggtca 480  
Qy 481 agtataacacatagcttggtagtagaagatgaatgaaattttcaccagctataca 540  
Dbb 481 agtataacacatagcttggtagtagaagatgaatgaaattttcaccagctataca 540  
Qy 541 gaacctattttcagttataaaacatgcccacatgcgaagtcocatgacatggtag 600  
Dbb 541 gaacctattttcagttataaaacatgcccacatgcgaagtcocatgacatggtag 600  
Qy 601 acctatgagctctatttattgtgaaggtgatacagtgctcagagcttttggatcgg 660  
Dbb 601 acctatgagctctatttattgtgaaggtgatacagtgctcagagcttttggatcgg 660  
Qy 661 ttttagtaaatctggtaccctgctcataagaattttaacaagcaagcatatgattggcaa 720  
Dbb 661 ttttagtaaatctggtaccctgctcataagaattttaacaagcaagcatatgattggcaa 720  
Qy 721 aggccttactgaagagacagctcaagctattgagccatatattaccaccttttttaac 780  
Dbb 721 aggccttactgaagagacagctcaagctattgagccatatattaccaccttttttaac 780  
Qy 781 aggttctgattgcttggaaaaacatctacagcatttgcagagcatgtcttgacttaa 840  
Dbb 781 aggttctgattgcttggaaaaacatctacagcatttgcagagcatgtcttgacttaa 840  
Qy 841 ttttggagctctacaattatgatagtcatttgcctcctctgttttaccacagcttgaat 900  
Dbb 841 ttttggagctctacaattatgatagtcatttgcctcctctgttttaccacagcttgaat 900  
Qy 901 ttaaaattaaagagcaaatgataatgagagcgccctacaagttgttaaaactactggcaaaa 960  
Dbb 901 ttaaaattaaagagcaaatgataatgagagcgccctacaagttgttaaaactactggcaaaa 960  
Qy 961 tgtttgggcaaaagattcagaaattgcttctcaaaaacagcactttgagctgact 1020  
Dbb 961 tgtttgggcaaaagattcagaaattgcttctcaaaaacagcactttgagctgact 1020  
Qy 1021 tgggaggttttaattgatccatgatgataacatccgcctgggaatgtgtgaaatttgcctagcc 1080  
Dbb 1021 tgggaggttttaattgatccatgatgataacatccgcctgggaatgtgtgaaatttgcctagcc 1080  
Qy 1081 attgtctcatgaacacatcctgatttagcaaaagacttaacagagctattcttaaaagtgaggt 1140  
Dbb 1081 attgtctcatgaacacatcctgatttagcaaaagacttaacagagctattcttaaaagtgaggt 1140

Qy 1141 cacatgacctgaggaagctatttagacatgatgttatttgttcaatattcacagctgcta 1200  
Dbb 1141 cacatgacctgaggaagctatttagacatgatgttatttgttcaatattcacagctgcta 1200  
Qy 1201 aaaaggaatttcttctggtcaatgatcacttacttaattttgtgagagagaaacattag 1260  
Dbb 1201 aaaaggaatttcttctggtcaatgatcacttacttaattttgtgagagagaaacattag 1260  
Qy 1261 aaaaacgatgagagtagcgaagaagccatgatgggacttgcacaaatttataagaata 1320  
Dbb 1261 aaaaacgatgagagtagcgaagaagccatgatgggacttgcacaaatttataagaata 1320  
Qy 1321 atgctttacagtcagcgtggaaaaagatgctcaaaacagatagcatggatcaaaagaca 1380  
Dbb 1321 atgctttacagtcagcgtggaaaaagatgctcaaaacagatagcatggatcaaaagaca 1380  
Qy 1381 aattgctacatatatatatacaaaatagatttgatgcgactacttgttgaacggatct 1440  
Dbb 1381 aattgctacatatatatatacaaaatagatttgatgcgactacttgttgaacggatct 1440  
Qy 1441 ttgctcaatcacatggttctctcaaaatttagaaactacagaacggatgaaatgcttatatt 1500  
Dbb 1441 ttgctcaatcacatggttctctcaaaatttagaaactacagaacggatgaaatgcttatatt 1500  
Qy 1501 acttgatgcccactggtttaaattgctgtgaagcattgaaatgtggaatgtc 1560  
Dbb 1501 acttgatgcccactggtttaaattgctgtgaagcattgaaatgtggaatgtc 1560  
Qy 1561 aaaaatctgctccgacatcaagtaagatttgcgtgacttgaattaaagcaaccccaaacag 1620  
Dbb 1561 aaaaatctgctccgacatcaagtaagatttgcgtgacttgaattaaagcaaccccaaacag 1620  
Qy 1621 atgcccagtgccaagggccatatattttcaaaagtgatggttattacaagaaatttagctgac 1680  
Dbb 1621 atgcccagtgccaagggccatatattttcaaaagtgatggttattacaagaaatttagctgac 1680  
Qy 1681 ctggttaaggtcaggtattcatgaagaaattcacacaggtgcttagaagatgatagaataa 1740  
Dbb 1681 ctggttaaggtcaggtattcatgaagaaattcacacaggtgcttagaagatgatagaataa 1740  
Qy 1741 taagaagcaggttagaagtagctttagtccaaacatgctcctgcaagcagctgaaagtt 1800  
Dbb 1741 taagaagcaggttagaagtagctttagtccaaacatgctcctgcaagcagctgaaagtt 1800  
Qy 1801 gtgtgcgtgaaataaactaaagatttgggcaaccccaaacagcctacaaatccttctcctgg 1860  
Dbb 1801 gtgtgcgtgaaataaactaaagatttgggcaaccccaaacagcctacaaatccttctcctgg 1860  
Qy 1861 aatgatacaagtttctctctgagagtagcacctgtgcacatagaccgaatctatca 1920  
Dbb 1861 aatgatacaagtttctctctgagagtagcacctgtgcacatagaccgaatctatca 1920  
Qy 1921 gtgctcttattaaacaagtgaaacaaatcaatagatgaacagcagatgataagatgagg 1980  
Dbb 1921 gtgctcttattaaacaagtgaaacaaatcaatagatgaacagcagatgataagatgagg 1980  
Qy 1981 gtgttcaactgatcaagccatcagagcaggtcttgaaactgcttaaggtactctcattta 2040  
Dbb 1981 gtgttcaactgatcaagccatcagagcaggtcttgaaactgcttaaggtactctcattta 2040  
Qy 2041 cacatcccatctcatttctctgtaaaacatttgaatcattactggtctgtctgaaaaa 2100  
Dbb 2041 cacatcccatctcatttctctgtaaaacatttgaatcattactggtctgtctgaaaaa 2100  
Qy 2101 tggatgataaaaaagtagcagaagctgacacaaattttcaaaaaacacaggaagcaaaa 2160  
Dbb 2101 tggatgataaaaaagtagcagaagctgacacaaattttcaaaaaacacaggaagcaaaa 2160  
Qy 2161 ttgaagaggattttccacacatcagatcagccttgccttcttcttcttcaatcacaaatcta 2220  
Dbb 2161 ttgaagaggattttccacacatcagatcagccttgccttcttcttcttcaatcacaaatcta 2220







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:28:02 ; Search time 520.53 seconds  
(without alignments)  
6358.268 Million cell updates/sec

Title: US-09-512-581-1

Perfect score: 5271

Sequence: 1 ccggagagcccgagtgag.....gtaaaaaaaaaaaaaa 5271

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	5271	100.0	5271	21	AAA28051
2	5271	100.0	5355	21	AAA28052
3	1112.8	21.1	3957	21	AAA47423
4	411.4	7.8	439	20	AAV87629
5	409.4	7.8	738	20	AAZ15259
6	364	6.9	405	16	AAT19926
7	333	6.3	351	21	AAZ42861
8	292	5.5	295	21	AAC02817
9	258.8	4.9	284	20	AAZ14459
10	253.8	4.8	1217	21	AAF18338
11	203.6	3.9	530	21	AAZ80598

12	165.6	3.1	936	22	AAF58252	Oligonucleotide D1
13	165.6	3.1	936	22	AAF58254	Oligonucleotide D1
14	165.6	3.1	936	22	AAF58257	Oligonucleotide D1
15	165.6	3.1	936	22	AAF58259	Oligonucleotide D2
16	165.6	3.1	936	22	AAF58262	Oligonucleotide D2
17	165.6	3.1	938	22	AAF58255	Oligonucleotide D1
18	163.4	3.1	936	22	AAF58252	Oligonucleotide D1
19	163.4	3.1	936	22	AAF58254	Oligonucleotide D1
20	163.4	3.1	936	22	AAF58257	Oligonucleotide D1
21	163.4	3.1	936	22	AAF58259	Oligonucleotide D2
22	163.4	3.1	936	22	AAF58262	Oligonucleotide D2
23	163.4	3.1	938	22	AAF58255	Oligonucleotide D1
24	149.6	2.8	150	19	AAI12049	Human biallelic po
25	149.6	2.8	150	19	AAI12050	Human biallelic po
26	122	2.3	161	21	AAC06611	Human secreted pro
27	69.4	1.3	244	22	AAF58238	Oligonucleotide D1
28	67.2	1.3	244	22	AAF58238	Oligonucleotide D1
29	62	1.2	6644	20	AAI33181	Base sequence of t
30	62	1.2	7372	20	AAI33182	Base sequence of t
31	62	1.2	7797	20	AAI33180	Cowpox virus bsr f
32	62	1.2	7996	20	AAI33184	Base sequence of t
33	58	1.1	6755	19	AAV21511	Staphylococcal bac
34	57.8	1.1	4000	18	AAT91902	Mannose-1-phosphat
35	55.2	1.0	5059	20	AAI84332	Stealth virus nucl
36	51.6	1.0	3399	17	AAT05868	Chicken leucocytos
37	51.4	1.0	72604	20	AAI10752	Genomic sequence o
38	50.8	1.0	1686	16	AAQ87587	DNA encoding Leuco
39	49.2	0.9	1935	21	AAC99088	Human pancreatic c
40	48.6	0.9	396	22	AAF94862	Human ovarian canc
41	48.2	0.9	1236	21	AAQ2163	Human colon cancer
42	48.2	0.9	10732	21	AAI10594	Gene encoding a su
43	48.2	0.9	49999	20	AAZ23895	Murine LORO homolo
44	47.6	0.9	6741	21	AAI10595	Gene encoding a su
45	47	0.9	6644	20	AAI33181	Base sequence of t

#### ALIGNMENTS

RESULT 1

AAA28051

ID AAA28051 standard; cDNA; 5271 BP.

XX

AC AAA28051;

XX

DT 01-DEC-2000 (first entry)

XX

DE Human androgen shutoff gene 3 (AS3) cDNA sequence.

XX

DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;

KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;

KW diagnosis; treatment; cytostatic; human; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT exon 1..46

FT /\*tag= a

FT /number= 1

FT exon 47..173

FT /\*tag= b

FT /number= 2

FT CDS 66..4241

FT /\*tag= c

FT /product= "AS3"

FT /note= "Androgen shutoff gene 3 protein, the CDS is

FT specifically claimed as SEQ ID #3"

FT exon 174..377

FT /\*tag= d

FT /number= 3

FT exon 378..464

FT /\*tag= e

FT /number= 4



**THIS PAGE BLANK (USPTO)**



```
QY 4513 gtggacagttggaccttactttgtggaccccatatcatttgggtcacatgcttttagccat 4572
Db 61 GTGGACAGTTGGACCTTACTTTGGTGACCCCATACGTTTGGTGCACATGCTTTAGCCAT 120
QY 4573 acacatggtaacattgactatgagcttctgtgaaagtgtaatgtgcgatggctatgtaga 4632
Db 121 ACACATGGTAACATTTGACTATGGAGTCTTGTGAAAGTGTAAATGTGGATGGCTATGTAGA 180
QY 4633 cataaagaagaactgttaaatctttcttttttttttttttttttttttttttttttttt 4692
Db 181 CATAAAGAAGAAGAACTGTAAATATCTTTTCTTTTTTAAATGTTCTGATTTCTGAAG 240
QY 4693 tgcttttatagctttttatctgcgcttttaaaactgacagtagcccgactgtttattggatct 4752
Db 241 TGCTTGTATAGCTTTTATCTGCGGCTTTAAACTGACAGTACCCGACTGTTTATTGGATCT 300
QY 4753 attgattgaaagaatttgttaggatagatctttaagcagtaactctgcaggtttgtat 4812
Db 301 ATTGATTGAAAGAAATTTGTAGATAGATCTTTAAGCAGTAATCTGTCAGTGTTCGTAT 360
QY 4813 ttgtatttctgcaatttactgtgaaaaaaatttgttttcaacaattgtgtcatttt 4872
Db 361 TTGTATCTCTGCAATTTTACTGTGAAAAAAATTTGTTTTCAACAAATGGTGCATTTT 420
QY 4873 cttgatgtcactatttgtggagagttaaatggctcttccctttgtgtatcttacctag 4932
Db 421 CTTGATGTCACATAATTGTTGGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACCTAG 480
QY 4933 tgttactcctgggaccccttaattctcagaggtgctaaattgtc 4977
Db 481 TGTNTACTCTGGGCACCCCTTAATCTTCAGAGGTGCTAAATTTGTC 525
```

Search completed: September 25, 2001, 16:19:16  
Job time: 6804 sec



```

RESULT 14
AI911784/c
LOCUS      535 bp      mRNA      EST      17-DEC-1999
DEFINITION      wd32h01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2329873 3', mRNA sequence.
ACCESSION      AI911784
VERSION        AI911784.1
KEYWORDS       AI911784.1 GI:5631639
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 535)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 708 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 446.
FEATURES             source
    Location/Qualifiers
        1..535
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2329873"
            /lab_host="Soares_NFL_T_GBC_S1"
            /lab_host="DH10B"
            /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
            a modified polylinker; Site_1: Not I; Site_2: Eco RI;
            Equal amounts of plasmid DNA from three normalized
            libraries (fetal lung NbHL19W, testis NHT, and B-cell
            NCI CGAP GCBI) were mixed, and ss circles were made in
            vitro. Following HAP purification, this DNA was used as
            tracer in a subtractive hybridization reaction. The driver
            was PCR-amplified cDNAs from pools of 5,000 clones made
            from the same 3 libraries. The pools consisted of
            I.M.A.G.E. clones 297480-302087, 682632-687239,
            726408-728711, and 729096-731399. Subtraction by Bento
            Soares and M. Fatima Bonaldo.
        205 a      90 c      81 g      159 t

BASE COUNT      205 a      90 c      81 g      159 t
ORIGIN
Query Match      10.0%; Score 528.6; DB 103; Length 535;
Best Local Similarity 99.3%; Pred. No. 5e-103;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4723 actgacagtcaccgactgtttattggtatctattgattgaaagaatttggtagataga 4782
|||||
Db 535 ACTGACAGTACCGAGCTGTTTATTGGACTATTGATTGAAAAAGAAATTCTTAGGATAGA 476

QY 4783 tcttaagcagtaactgtcagtggttattgtatttctgcaatttctactgtgaaaaa 4842
|||||
Db 475 TCTTAAGCAGTAATCTGTCTGAGTGTGTTGTTATTTCTCTGCAATTTACTGTGAAAAA 416

QY 4843 aaattgttttcaacaatttggtgtcatttctgtgatgtcactatttggagagttaa 4902
|||||
Db 415 AAATTTGTTTTTACAAATTTGGTGTCATTTCTTGATGTCACATATTTGTTGGAGAGTTAA 356

QY 4903 tggctcttcccttctgtgtatcttactagtgttactcctgggcacccttaattcttcag 4962
|||||
Db 355 TGGTCTCTCTCCCTTGTGTATCTTACTAGTGTGTTTACTCCTGGCACCCCTTAATCTTCAG 296

QY 4963 aggtgctaataatgtctgcattacacagcagagatgcctctgatagagacacaccatgc 5022
|||||
Db 295 AGGTGCTAAATTTGTCGTCCATTACACAGAGGATGCCTCTGATAGGAGGACCAACCATGC 236

QY 5023 aaattgtgaatagtcctgaagttcttgattacttctacacctcagattgttctcc 5082
|||||

```

---

```

Db 235 AAATTGCTGAANTAGTCCTCCTGAAGTTCCTGGATTACTTTACACCTCAGTATTGATTGTCC 176
QY 5083 agaattttctggcctttcatcggaatgaaattttaagaagaagaagatttaaaagtatttta 5142
|||||
Db 175 AGAATTTTCTGGCCTTTCATGGCAATGAAATTTTAAAGAAGAAGATTTTAAAGTATTTTA 116
QY 5143 attttaagaagtgctgtataaaaaataatgtactgaattcttcttccattttatccatt 5202
|||||
Db 115 ATTTTAAAGAGTGTGTATATAAATAATGACTGACTGAATTTCTTTATCCCATTTTATCATCCT 56
QY 5203 tcagtttttataatctactgtatcatcaataaaattctgttaattggaatgaatgaataaa 5257
|||||
Db 55 TCAGTTTTTATTAACTACTGTATCAATAAATTTCTGTAATTTTAAAGAGAAAAA 1

RESULT 15
BF111072      525 bp      mRNA      EST      20-OCT-2000
LOCUS      7n43b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3567291 3',
DEFINITION      mRNA sequence.
ACCESSION      BF111072
VERSION        BF111072
KEYWORDS       BF111072.1 GI:10940762
SOURCE         EST.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 525)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 470.
FEATURES             source
    Location/Qualifiers
        1..525
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3567291"
            /tissue_type="carcinoid"
            /lab_host="DH10B"
            /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
            modified polylinker; Plasmid DNA from the normalized
            library NCI_CGAP_Lu5 was prepared, and ss circles were
            made in vitro. Following HAP purification, this DNA was
            used as tracer in a subtractive hybridization reaction.
            The driver was PCR-amplified cDNAs from a pool of 5,000
            clones made from the same library (cloneIDs
            1414920-1417991 and 1520904-1522439). Subtraction by Bento
            Soares and M. Fatima Bonaldo.
        133 a      81 c      109 g      201 t

BASE COUNT      133 a      81 c      109 g      201 t
ORIGIN
Query Match      9.9%; Score 520.8; DB 144; Length 525;
Best Local Similarity 99.4%; Pred. No. 2.3e-101;
Matches 522; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4453 ttttggctcaagcttgaggctgaataaaagcccttctgtatgcacaaaatgggactgtgaaga 4512
|||||
Db 1 TTTTGGCTCAAGCTTGAGGCTGAATAAAGCCCTTGATGTCACAAAATGGGACTGCTGAAGA 60

```



```

Best Local Similarity 99.3%; Pred. No. 1.6e-104;
Matches 538; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4715 ggcattaaactgacagcaccgactgtttattgagctattgattgaaagaatttggtt 4774
Db 542 GGCCTTTAAATGACAGTACCAGCTGTTTATGAGCTATATGATTTGAAAAGAAATTTGTT 483
QY 4775 aggtatgcttaagcagtaactctgctgagttgttattgtatttctgcaattttact 4834
Db 482 AGGATAGATCTTAAGCAGTAATCTGTCAGTGTTTGTATTTGATCTCTCAATTTTACT 423
QY 4835 gtgaaaaaaattgttttcaacaattggtgtcatttcttctgagtcactatttgggga 4894
Db 422 GTGAAAAAAATTTGTTTCAACAATTTGGTCTCATTTTCTTGATGTCACATTTTGTGGA 363
QY 4895 gagttaaattggtctctccctcttctgtgtatcttaactagtggtttactctgggcaacctta 4954
Db 362 GAGTTAAATGCTCTCTTCCCTTTTGTGTATCTTACTAGTGTGTACTCTGGCACCCTTA 303
QY 4955 atcttcagagtgctaaattgtctgcccattacaccagagagatgctctgataggagac 5014
Db 302 ATCTTCAGAGTGCTAAATTTGTCGCCATTACACCAGAGGATGCCCTCTGATAGGAGAC 243
QY 5015 aaccatgcaaatgtgaaatagctgctgaagtcttcttgattactttacacctcagattga 5074
Db 242 AACCATGCAAAATTTGAAATAGTCTCTCAAGTCTCTTGATTTACTTTACACCTCAGTATTGA 183
QY 5075 ttgtcccaagaattttctgccccttcagcaatgaaattttaaagaagaagattttaa 5134
Db 182 TTTGTCCCAAAATTTCTGCCCCTTTTCAAGCAATTTTAAAGAAAGAAATTTTAAA 123
QY 5135 gtattttaaattaaagagtggttataaaatgactgaattcttaccatttta 5194
Db 122 GTATTTTAAATTTAAAGAGTGCTGTATTAATAATATGACTGAATTTCTTATCCCATTTTA 63
QY 5195 tcattccttcagtttttatttaatactactgtatcaataaaattctgtaatttgaatgagta 5254
Db 62 TCATCCTTTCAAGTTTATTATTAATCCCTGTATCAATAAAATTTCTGTAATTTGAATGAGAA 3
QY 5255 aa 5256
Db 2 AA 1

RESULT 13
BF142489 944 bp mRNA EST 24-OCT-2000
LOCUS 601788573F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016155 5',
DEFINITION mRNA sequence.
ACCESSION BF142489
VERSION BF142489.1 GI:10981529
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 944)
NTL-MGC http://mgl.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM5263 row: a column: 20
High quality sequence stop: 620.
Location/Qualifiers
1. 944
FEATURES
source

```

```

/organism="Mus musculus"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:4016155"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site: 1. Notl;
Site_2: SalI; transgenic model WNT-1, expression driven by
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dr. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 277 a 221 c 218 g 228 t
ORIGIN
Query Match 10.1%; Score 532.8; DB 145; Length 944;
Best Local Similarity 86.7%; Pred. No. 6.8e-104;
Matches 633; Conservative 0; Mismatches 92; Indels 5; Gaps 4;

QY 2557 tgatggttcgactggtcttggaataataatcacagtaaatcacaggaacttctacct 2616
Db 1 TGATGGTTCGATGGCTACTTTGGAATGAAGATAATACAGTAAGTCAGGAACATTTCCACCC 60
QY 2617 taagattgctaaacaataattgcatagtgatgagacttgacagagaggggaaatta 2676
Db 61 TCAGATTACTAACACAGATACTGCTAGTAGTGGGATTTGACAGAACACAGGGAATTA 120
QY 2677 gtaaacagatatgtcactgtgagactgtctgtggagtgctattgtgaagctggcac 2736
Db 121 GTRAACAGATATGTCACGGCTGAGACTTGTCTGGGAGTGCCTATTGTGAAGCTGGCAC 180
QY 2737 aagaacctgttaccatgaaatcacatcattagaaataatcacatcagctatgtcattagta 2796
Db 181 AGGACCCCTGTATACAGAGATCATACACTGGAGCATACACGCTGTGTGCATTTAGCCA 240
QY 2797 tcaacgatgaatgctacaagtaagacaagtgttggccagaaacttcacaaagccctt 2856
Db 241 TCATGATGAGTGTATCAAGTCAGGAGGTGTGCGTTCAGAAATTCACAAAGCCCTTT 300
QY 2857 ccggtttacggttccacttgagtatatggcaatctgtgccctttgtgcaaaagactcgt 2916
Db 301 CCCGCTTACGGCTTCCCTTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCCTG 360
QY 2917 taaaggagagaagatcagctaggaatgtttggggaataataataatgaagggcggg 2976
Db 361 TGAAGAGAGAGCGGAGCCCATGCTAGACAGTGTGGTGAAGAACATCCTGTGAGGAGGG 420
QY 2977 agtatctgaagcagcagctgctgttagtgaataattattgtctctctaccagagtatg 3036
Db 421 AGTACCTGAAGCAGCATGCAGCTGTTAGTGAATAATTTATTGTCTCTTCTACCAAGATG 480
QY 3037 ttgttccatatacaattcaacttttggcacatgaccacagattattgtcaaaagtacaggata 3096
Db 481 TGGTTCATATACAAATTCACCTTTGGGCACATGACCCAGATTTATGCAACGTACAGGATA 540
QY 3097 ttgaacaacttaagaatgttaagaatgtcttggtttcttctggaataataatagcta 3156
Db 541 TTGAACAACTTAAAGATGTGACAAGATGCTCTGGGTTGTTTCTGGAGATATTGATGGCTA 599
QY 3157 aaaaataaacaagtcacgcttttcatcagaagaatggtagaataattataaacaacaa 3216
Db 600 AAAATGAACAAACAG-CATGCTTTTATCAGAAAATGGTAGAAAATATTAAACAGACCA 658
QY 3217 aagatgccccagaccagatgagtcacaaataatgaataaaactgtacactgtgtgtgatg 3276
Db 659 AGATTGCTCCAGGACCGCATGAT-CACCACTGACTGCACAACTGT--CCCCGTGTGTGATG 715
QY 3277 ttgcatgaa 3286
Db 716 TTGCCATGGA 725

```



```

Db 481 CAGATGATGCAGAAATGAATGAAAACTGTACACTGTGTGTGATGNGCCATGAATATCA 540
|||||
QY 3292 tcagtcaagaagtactacatac-aggttggaatctcc 3328
|||||
Db 541 TCATGTCAGAAAGACTACTACATACNAGCTTGGATCTCC 578
|||||

RESULT 11
BF002153/c
LOCUS 7h23a12.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:3316798 3',
DEFINITION mRNA sequence.
ACCESSION BF002153
VERSION BF002153.1 GI:10702428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
FEATURES
source
location/Qualifiers
1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3316798"
/clone_lib="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 209 a 93 c 85 g 153 t
ORIGIN
Query Match 10.2%; Score 536.8; DB 143; Length 540;
Best Local Similarity 99.6%; Pred. No. 8.6e-105;
Matches 538; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4711 ctgcggcttaactgacagtagcccgactgtttattggatctattgattgaaagaatt 4770
|||||
Db 540 CTGCGCGCTTAAACTGACAGTACCCGACGTGTTTATGGATCTATTGATTTGAAAGAATTT 481
|||||
QY 4771 tcttaagatagatcttaagcagtaactctcagtggtttgtattgtatttctgcaattt 4830
|||||
Db 480 TGTAGATAGATCTTAAAGCAGTAATCTGTACGTGTTGTATTGTTGTTCTCTGCAATTT 421
|||||
QY 4831 tactgtgaaaaaaattgttttcaacaattggtgtcattttcttctgattgtcactattgt 4890
|||||
Db 420 TACTGTGAAAAAAATTTGTTTCACAAATGCTGTGTCATTCTTCTGATGTCACATTATTGT 361
|||||

```

```

QY 4891 tggagagttaaatggtctctcttccctttgtgtatcttaccatggttactcctggcacc 4950
|||||
Db 360 TGGAGAGTTAAATGGTCTCTCTCCCTTTGTGTATCTTACCTAGTGTCTTACCTCTGGGCACC 301
|||||
QY 4951 cttaatcttcagagtgctaaattgtctgccattacacagaagaagatgcctctgatagga 5010
|||||
Db 300 CTTAATCTTCAGAGGTGCTAAATGTCTGCCATTACACCAGAAGGATGCCCTCTGATAGGA 241
|||||
QY 5011 ggacaaccatgcaaatgtgaaatagtcctggaagtcttgattcttaccatcactcagta 5070
|||||
Db 240 GGACAACCATGCAAAATGTGAATAGTCTGGAAGTTCTTGGATTACTTTACACCTCAGTA 181
|||||
QY 5071 ttagattgtccagaattttctggtcccttccatggcaatgaaatttttaagaagaagatt 5130
|||||
Db 180 TTGATTGTGCCCAGAAATTTCTGGCCTTTCATGGCAATGAAAAATTTTAAAGAAGAAATT 121
|||||
QY 5131 taaagtatttttaatttaagagtggtgtatataaaataatgactgaattcttattcccat 5190
|||||
Db 120 TAAAGTATTTTAAATTTAAAGAGTGTGTATATAAAATGATGACTGAATCTTTATCCCAT 61
|||||
QY 5191 ttatcatcctttcagtttttatttaatactactgtatcaataaaattctgttaattgaatg 5250
|||||
Db 60 TTTATCATCTTTCAGCTTTTATTATTAATCCACTGTATCAATAAAATCTGTAAATTGAATG 1
|||||

```

```

RESULT 12
LOCUS AW303630/c
DEFINITION XV21e03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2813788 3', mRNA sequence.
ACCESSION AW303630
VERSION AW303630.1 GI:6713319
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
FEATURES
source
location/Qualifiers
1..542
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2813788"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 208 a 93 c 84 g 157 t
ORIGIN
Query Match 10.2%; Score 535.6; DB 114; Length 542;

```















QY 2792 agctataacagatgaatgtctatcaagtaagacaaagtgttggccagaaacattcacaaagg 2851  
 Db 480 AGCTATCAAGGATGAATGTATCAAGTAAGACAAGTGTGGCCAGAAAATTCACAAAGG 539  
 QY 2852 cctttccctttac-ggcctccacttgagtatatgcaatctgtgcccctttgtgcaa-aa 2909  
 Db 540 CTTTCCCGTTTACGGGCTTCCACTTGGATATATGCAATCTGTGCCCTTGGTGCAAGAA 599  
 QY 2910 gatcctgtataa---ggagagaagagctcatgctaggtgcaatgtttgtgaaaaataataa 2966  
 Db 600 GATCTGCTTAAGAGAGAGAAAGAGCTCATGCTAGGCAATGTTGGTGAACAATATAAAT 659  
 QY 2967 -gtagggcgggagtagtctgaaagacagatgcagctgttagtgaataattattgtctcttc 3024  
 Db 660 TGTCAAGGGGGAGGATCTCTTGAAGCAGATTTGGGCTGTAGTGAGAAACTTATGGGCTTC 719  
 QY 3025 taccagagtagtgtttccatatacaattcaccttccctttgacacatgacc 3073  
 Db 720 TACCACAG-ATTTGGTGCCCTATATGATTCACCTTTTGGGGCCCTTGGCCC 767

RESULT 5  
 BG432960 729 bp mRNA EST 14-MAR-2001  
 LOCUS 602495783F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4609533 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG432960  
 VERSION BG432960  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 729)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 plate: LLCM1353 row: e column: 22  
 High quality sequence stop: 722.  
 Location/Qualifiers  
 1. .729  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4609533"  
 /clone\_lib="NIH\_MGC\_75"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggcgcctcgcc); Site\_2: SfiI (ggcattatggcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCCCATATGGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.65  
 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH\_MGC Library."

## FEATURES

source  
 BASE COUNT 225 a 129 c 144 g 231 t  
 ORIGIN  
 Query Match 11.6%; Score 612.4; DB 153; Length 729;  
 Best Local Similarity 97.9%; Pred. No. 5.4e-121;  
 Matches 715; Conservative 0; Mismatches 6; Indels 9; Gaps 9;

QY 428 gagccccacaattcaataggtatttttttacttgagaacattgcttgggtcgaagtcata 487  
 Db 1 GAGCCCCACAGTTCATTAAGTATTTTATTTACTTGAGAACA-TGCTTGGGTCAAGTCATA 59  
 QY 488 taacatatgctttgagttagaagatagcaatgaaattttcacccagctatacacaaacctt 547  
 Db 60 TAACATATGCTTTTGAGTTAGAAATAGCAATGAAATTTTCCACCAGCTATACAGAACCTT 119  
 QY 548 attttcagttataaacaatggccacacatcagaagtcacatatgcacatggtagacacctat 607  
 Db 120 ATTTTCAGTTATTAACAATGGCCACCAATCAGAAAGTCCATATGCACATGTAGACCTTAT 179  
 QY 608 gagctctattatttgaagggtgatacagtgctcagggagctttttggatagcaggttttagt 667  
 Db 180 GAGCTCTATTATTGTGAAGGTGATACAGTGTCTCAGAGAGCTTTTGGATACGGTTTTAGT 239  
 QY 668 aaatctgggtacctgtctcataagaatttaaacagaagcagcatatgatttggcacaaggcttt 727  
 Db 240 AAATCTGGTACCTGCTCATTAAGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGGCTTT 299  
 QY 728 actgaagagagacagctcaagctatttgagccatatattaccacatttttttaatacagggtctt 787  
 Db 300 ACTGAAGAGAGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTAAATCAGGTTCT 359  
 QY 788 gatgcttgggaaaaacatctatcagcgatattgtcagagcatgtc-tttgacttaattttg 846  
 Db 360 GATGCTTGGGAAAACATCTATCAGCGATTTTGTGAGCATGTCTTTTGACTTAATTTTGG 419  
 QY 847 agctctacaatatgtatgctcatttgcctctctctgttttaccacagc-ttgaatttaaa 905  
 Db 420 AGCTCTACAAATATTGATGATGATTTGCTGCTCTCTGTTTACCCACAGCTTTGAATTTAA 479  
 QY 906 ttaaaagagcaatgataatgaggagcgcctacaagtggttaaaactactggcaaaaattgtt 965  
 Db 480 TTAAGAGCAATGATAATGAGGAGCGCTCAAGTGTGTTAAACTACTGTGCAAAAATGCTT 539  
 QY 966 ggggcaaggattcagaattggtcttc-aaacaagccacctttggcagtgctac-ttgg 1023  
 Db 540 GGGGCAAGGATTCAGAAATTTGGCTTCTCAAAAACAAGCCACTTTGGCAGTGCTACTTTGG 599  
 QY 1024 gcaggtttaaataatgccatgccatgtaccac-tccgcctggaatgtgga-aatttgcagcca 1081  
 Db 600 GCAGGTTTAATGATATCCATGTACCAATTCGCCCTGGAATGTGTGACCATTGTGTCACCA 659  
 QY 1082 ttgtctcagaaccatcctctgatttagc-aaaagacttaacagagatctctt-aaagtgaag 1139  
 Db 660 TTGCTCATGAACCATCCTGATTTAGCAAAAAGATTTTAACAGAGATCTCTTAAAAAGTCAGG 719  
 QY 1140 tcacatgacc 1149  
 Db 720 TCACCTTGAC 729

RESULT 6  
 BE504550 596 bp mRNA EST 04-AUG-2000  
 LOCUS hz58a05.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3212144 3',  
 DEFINITION similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA  
 sequence.  
 ACCESSION BE504550  
 VERSION BE504550.1 GI:9706958  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 596)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)



**COMMENT**

Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 317

See also: **FAITH**.

seq primer: forward.  
Location/Qualifiers  
1: .648

```
/organism="Homo sapiens"
```

/db\_xref="taxon:9606"

```
/clone_lib="MAGE resequences,
```

```
257 0 112 0 04 0 105 4
/note="Vector: pBluescriptS
```

BASE C  
ORIGIN

Query Match	12.0%	Score 632	DB 122	Length 648
-------------	-------	-----------	--------	------------

Best Local Similarity 98.5%; Pred. No. 3.3e-125;

Matches 638; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Ov 4604 gaaagtgtaatgttcgatacctatgtagacataaagaaagaaacttgtataatattctttttt. 4663

\_\_\_\_\_

Db 648 GAAAGTGAATGTCGAATGGCTATGTAGACATAAAGAAGAACTTGTAATAATCTTTTC 589

Qy 4664 ctttttttaagtgttctgatttctgaagtgcttgatagcttttatctgcggctttaa 4723

[illegible]

DD 388 TTTTAAAGTTCTGATTCIGGAAGTGCTGTGTA TAGCTTTTATCTGGGCTTTAAA 329

ov 4724 ctgacagtaccgcgactgtttatttggatctattgatttgaagaagaattttattaggatagat 4783

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 528 CTGACAGTACCCGACTGTTATTGGATCTATTGATTGAAAAGAAATTCTTAGGATAGAT 469

Qy 4784 cttaacgagtaattctgtcagtggttgattgtattttctgcaattttactgtgaaaaa 4843

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

DD 466 CTTAAGCAGTAATCTGTCTAGTGTTTGTATTTGTATTCCTCTGC AATTTTACTGTGAAAAA 409

Oy 4844 aatttgttttcaacaatttgtgtcattttcttgatctcactatttcttttgagagattaaat 4903

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 408 AATTGTTTCAACAATTGGTGTCTTCTTGATGTCACCTATTGTTGGAGAGTTAAAT 349

Qy 4904 ggtctcttcccttgtgtatcttaccttagtgtttactcctgggcaccttaattcaga 4963

240

DD 348 GGTCCTTGTTGTATCTACCTAGTGTTACTCCGGGCACCTTAATCTTCAGA 289

Qy 4964 ggttgttaatttctctgccattacaccagaaagatgccctctgatataggaagacaaccatgca 5023

## 4 RESULT

BE873840

LOCUS	BE873840	848 bp	mRNA	EST	20-OCT-2000

DEFINITION 601483992F1 NI

mRNA seq







## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA-yes.

## FEATURES

source

Location/Qualifiers

1..791

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3086220"

/clone\_lib="NCI-CGAP\_Sub8"

/lab\_host="DHI0B (Life Technologies)"

/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; NCI-CGAP\_Sub8 is a subtracted library derived from NCI-CGAP\_Sub5. The NCI-CGAP\_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI-CGAP\_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI-CGAP\_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population); a pool of clones from NCI-CGAP\_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population); NCI-CGAP\_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI-CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550); 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG\_LIB=NCI-CGAP\_Lu5

TAG\_TISSUE=lung

TAG\_SEQ=CAAC

BASE COUNT 270 a 151 c 144 g 226 t

## ORIGIN

Query Match 13.6%; Score 716; DB 149; Length 791;  
Best Local Similarity 99.6%; Pred. No. 3, le-143;  
Matches 749; Conservative 0; Mismatches 0; Indels 3; Gaps 3;  
Qy 2745 tttaccatgaatcaccattagacaataatcagctatgtgcattagctatcaacgat 2804  
Db 13 TGTTCACATGAATCATCATATAGCAATATACAGTATGTGCATTAGCTATCAACGAT 72  
Qy 2805 gaatgctatgaagaagaagttgtcccaagaactccaaaggccttccggttta 2864  
Db 73 GAATGCTATCAAGTAAGCAAGTGTTCGCCAGAACTTCACAAAGGCTTTCCGCTTTA 132  
Qy 2865 cggctccacttgatgatgcaatctgtgcccttggcaaaagatcctgtaagagag 2924  
Db 133 CGGCTTCACATGAGTATATGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAGGAG 192  
Qy 2925 agaagagctcatgctaggaagaatgtttggtaaaaaataataatgaagcggtgatcttg 2984  
Db 193 AGAAGAGCTCATGCTAGGCAATGTTTGGTGAATAATATAAATGTAAGCGGGAGTATCTG 252  
Qy 2985 aagcagcagctgagctgttagtgaaaaaattattgtctctaccagagatattgtttcca 3044  
Db 253 AAGCAGCATGAGCTGTTAGTGAATAATATTGTCTCTCTACAGAGTATGTTGTTCCA 312  
Qy 3045 tatacaattcacttttggcacatgaccacagattatgtcaagtagcagagattgaacaa 3104  
Db 313 TATACAAATTCACCTTTTGGCAGCATGACCAGATTAATGCAAAAGTACAGGATATTGAACAA 372  
Qy 3105 cttaagaagtgtaagaagatgtctttgttctgttgaaataattatggtgtaaaatgaa 3164  
|||||

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	716	13.6	791	149	BF509252	UI-H-BI4-BF509252
2	663.4	12.6	734	175	BG256731	602370939
3	632	12.0	648	122	AW973437	EST385430
4	621.2	11.8	848	141	BE873840	601483992
5	612.4	11.6	729	153	BG432960	602495783
6	596	11.3	596	136	BE504550	hz58a05.x
7	591.4	11.2	802	153	BG387646	602412496
8	583.6	11.1	738	122	AW976150	EST388259
9	560.4	10.6	568	118	AW629426	hi57b12.x
10	559.2	10.6	578	120	AW770571	hl86f11.x
11	536.8	10.2	540	143	BF002153	7h23a12.x
12	535.6	10.2	542	114	AW030630	xv21e03.x
13	532.8	10.1	944	145	BF142489	601788573
14	528.6	10.0	935	103	AI911784	wd32h01.x
15	520.8	9.9	525	144	BF111072	7n43b02.x
16	503.8	9.6	608	32	AV685306	AV685306
17	497.4	9.4	608	149	BF471938	UI-M-BH3-BF471938
18	495	9.4	683	136	BE532986	601235013
19	482.8	9.2	486	23	AI681422	tx46e01.x
20	482.6	9.2	535	137	BE551003	7b67b03.x
21	479.4	9.1	888	155	BG571128	602591748
22	478.4	9.1	480	23	AI655429	ts98f06.x
23	474.8	9.0	490	149	BF511856	UI-H-BI4-BF511856
24	467.4	8.9	469	9	AA563884	nk19e08.s
25	466.8	8.9	470	4	AA282190	zs99b08.r
26	456.2	8.7	461	15	AI016896	ou31c08.x
27	437.2	8.3	442	7	AA447433	zw93f10.x
28	430.4	8.2	433	18	AI275866	q168c06.x
29	423	8.0	423	23	AI680124	tw65b01.x
30	407.4	7.7	409	9	AA625961	zu92e08.s
31	406.8	7.7	410	14	AA987361	or92g04.s
32	406.8	7.7	648	137	BE570606	601329907
33	402.8	7.6	406	18	AI299646	qnl2a10.x
34	393	7.5	405	16	AI140790	ocl5b04.x
35	392	7.4	434	116	AW483768	56376 MAR
36	389	7.4	429	145	BF189753	235446 MA
37	388.4	7.4	390	148	BF446410	7p36h09.x
38	388.4	7.4	391	10	AA630616	ac11e07.s
39	379.2	7.2	735	154	BG532020	602561049
40	377	7.2	445	142	BE948523	UI-M-BH3-BE948523
41	376.4	7.1	738	105	AL045878	DRFp434J
42	373.4	7.1	580	149	BF451878	uz80g04.y
43	373.4	7.1	738	32	AV710229	AV710229
44	368.2	7.0	595	143	BF043498	BP250017A
45	365	6.9	703	112	AW175091	fi32b08.y

## ALIGNMENTS

RESULT 1  
LOCUS BF509252  
DEFINITION UI-H-BI4-aow-c-07-0-UI.s1 NCI-CGAP\_Sub8 Homo sapiens cDNA clone  
IMAGE:3086220 3', mRNA sequence.  
ACCESSION BF509252  
VERSION BF509252.1 GI:11592550  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 791)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)



117: gb\_est48: \*  
118: gb\_est49: \*  
119: gb\_est50: \*  
120: gb\_est51: \*  
121: gb\_est52: \*  
122: gb\_est53: \*  
123: gb\_est54: \*  
124: gb\_est55: \*  
125: gb\_est56: \*  
126: gb\_est57: \*  
127: gb\_est58: \*  
128: gb\_est59: \*  
129: gb\_est60: \*  
130: gb\_est61: \*  
131: gb\_est62: \*  
132: gb\_est63: \*  
133: gb\_est64: \*  
134: gb\_est65: \*  
135: gb\_est66: \*  
136: gb\_est67: \*  
137: gb\_est68: \*  
138: gb\_est69: \*  
139: gb\_est70: \*  
140: gb\_est71: \*  
141: gb\_est72: \*  
142: gb\_est73: \*  
143: gb\_est74: \*  
144: gb\_est75: \*  
145: gb\_est76: \*  
146: gb\_est77: \*  
147: gb\_est78: \*  
148: gb\_est79: \*  
149: gb\_est80: \*  
150: gb\_est81: \*  
151: gb\_est82: \*  
152: gb\_est83: \*  
153: gb\_est84: \*  
154: gb\_est85: \*  
155: gb\_est86: \*  
156: gb\_est87: \*  
157: gb\_est88: \*  
158: gb\_est89: \*  
159: gb\_est90: \*  
160: gb\_est91: \*  
161: gb\_est92: \*  
162: gb\_est93: \*  
163: gb\_est94: \*  
164: gb\_est95: \*  
165: gb\_est96: \*  
166: gb\_est97: \*  
167: gb\_est98: \*  
168: gb\_est99: \*  
169: gb\_est100: \*  
170: gb\_est101: \*  
171: gb\_est102: \*  
172: gb\_est103: \*  
173: gb\_est104: \*  
174: gb\_est105: \*  
175: gb\_est106: \*  
176: gb\_est107: \*  
177: gb\_est108: \*  
178: gb\_est109: \*  
179: gb\_est110: \*  
180: gb\_est111: \*  
181: gb\_est112: \*  
182: gb\_est113: \*  
183: gb\_est114: \*  
184: gb\_est115: \*  
185: gb\_est116: \*  
186: gb\_est117: \*  
187: gb\_est118: \*  
188: gb\_est119: \*  
189: gb\_est120: \*

190: gb\_est121: \*  
191: gb\_est122: \*  
192: gb\_est123: \*  
193: gb\_est124: \*  
194: gb\_est125: \*  
195: gb\_est126: \*  
196: gb\_est127: \*  
197: gb\_est128: \*  
198: gb\_est129: \*  
199: gb\_est130: \*  
200: gb\_est131: \*  
201: gb\_est132: \*  
202: gb\_est133: \*  
203: gb\_est134: \*  
204: gb\_est135: \*  
205: gb\_est136: \*  
206: gb\_est137: \*  
207: gb\_est138: \*  
208: gb\_est139: \*  
209: gb\_est140: \*  
210: gb\_est141: \*  
211: gb\_est142: \*  
212: gb\_est143: \*  
213: gb\_est144: \*  
214: gb\_est145: \*  
215: gb\_est146: \*  
216: gb\_est147: \*  
217: gb\_est148: \*  
218: gb\_est149: \*  
219: gb\_est150: \*  
220: gb\_est151: \*  
221: gb\_est152: \*  
222: gb\_est153: \*  
223: gb\_est154: \*  
224: gb\_est155: \*  
225: gb\_est156: \*  
226: gb\_est157: \*  
227: gb\_est158: \*  
228: gb\_est159: \*  
229: gb\_est160: \*  
230: gb\_est161: \*  
231: gb\_est162: \*  
232: gb\_est163: \*  
233: gb\_est164: \*  
234: gb\_est165: \*  
235: gb\_est166: \*  
236: gb\_est167: \*  
237: gb\_est168: \*  
238: gb\_est169: \*  
239: gb\_est170: \*  
240: gb\_est171: \*  
241: gb\_est172: \*  
242: gb\_est173: \*  
243: gb\_est174: \*  
244: gb\_est175: \*  
245: gb\_est176: \*  
246: gb\_est177: \*  
247: gb\_est178: \*  
248: gb\_est179: \*  
249: gb\_est180: \*  
250: gb\_est181: \*  
251: gb\_est182: \*  
252: gb\_est183: \*  
253: gb\_est184: \*  
254: gb\_est185: \*  
255: gb\_est186: \*  
256: gb\_est187: \*  
257: gb\_est188: \*  
258: gb\_est189: \*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:25:52 ; Search time 6640.39 Seconds  
(without alignments)  
7503.474 Million cell updates/sec

Title: US-09-512-581-1  
Perfect score: 5271  
Sequence: 1 ccggagagcccgagtgag.....gtacaaaaaaaaaaaaa 5271

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*  
48: gb\_est48.\*  
49: gb\_est49.\*  
50: gb\_est50.\*  
51: gb\_est51.\*  
52: gb\_est52.\*  
53: gb\_est53.\*  
54: gb\_est54.\*  
55: gb\_est55.\*  
56: gb\_est56.\*  
57: gb\_est57.\*  
58: gb\_est58.\*  
59: gb\_est59.\*  
60: gb\_est60.\*  
61: gb\_est61.\*  
62: gb\_est62.\*  
63: gb\_est63.\*  
64: gb\_est64.\*  
65: gb\_est65.\*  
66: gb\_est66.\*  
67: gb\_est67.\*  
68: gb\_est68.\*  
69: gb\_est69.\*  
70: gb\_est70.\*  
71: gb\_est71.\*  
72: gb\_est72.\*  
73: gb\_est73.\*  
74: gb\_est74.\*  
75: gb\_est75.\*  
76: gb\_est76.\*  
77: gb\_est77.\*  
78: gb\_est78.\*  
79: gb\_est79.\*  
80: gb\_est80.\*  
81: gb\_est81.\*  
82: gb\_est82.\*  
83: gb\_est83.\*  
84: gb\_est84.\*  
85: gb\_est85.\*  
86: gb\_est86.\*  
87: gb\_est87.\*  
88: gb\_est88.\*  
89: gb\_est89.\*  
90: gb\_est90.\*  
91: gb\_est91.\*  
92: gb\_est92.\*  
93: gb\_est93.\*  
94: gb\_est94.\*  
95: gb\_est95.\*  
96: gb\_est96.\*  
97: gb\_est97.\*  
98: gb\_est98.\*  
99: gb\_est99.\*  
100: gb\_est100.\*  
101: gb\_est101.\*  
102: gb\_est102.\*  
103: gb\_est103.\*  
104: gb\_est104.\*  
105: gb\_est105.\*  
106: gb\_est106.\*  
107: gb\_est107.\*  
108: gb\_est108.\*  
109: gb\_est109.\*  
110: gb\_est110.\*  
111: gb\_est111.\*  
112: gb\_est112.\*  
113: gb\_est113.\*  
114: gb\_est114.\*  
115: gb\_est115.\*  
116: gb\_est116.\*











```
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.68; Score 187.5; DB 21; Length 1562;
Best Local Similarity 18.8%; Pred. No. 0.0009;
Matches 304; Conservative 261; Mismatches 584; Indels 471; Gaps 77;

QY 22 ISDKISK-----PEMVRRLKMWVK-----TFMDMDQDSEEEKELYLNALHLASDFLKL 70
DB 20 ispnkskwqlentdmkskvqkaeaslaqfdqnleiklsnl-----esevs 69
QY 71 HPGKDVRLLVACCLADIPRIYAPEAPYTPDKLKDIFMFIITROLKGLDTRKSPQFNRYFY 130
DB 70 ragedsrvlle-----ratraeeavet-----lreslskvekevessllqyqq 112
QY 131 LLENIAWKSINICFELEDSNEIFITQLYRTLFVSINNGHNOKVHMHVVDLMSIICEGDT 190
DB 113 clqnia-----dledrislaqkeagevderanraaeatlalkqslvssetdkeaal 163
QY 191 VS-QELDLTVLVNLPVAPHNKLNQAYDLAKALLKRTAQAIPIYITTFNQVLMGLGKTSIS 249
DB 164 vyyqqlklti-snl---eerlhkaeed---srlnqraenaagevesl-----kqkvs 209
QY 250 DLSEHFVOLIILEYNIDSHLLSVLPQLEFKL-KSNDNEERLQ-----VVKL----- 295
DB 210 kliee-----neayelqyqcltdiadlklklfhaqetgrlsieidgvaklkfaeakc 264
QY 296 -----LAKMFGAKDSELASQNKP---LWQCY----- 323
DB 265 vvlersnqlhseidgllekignqsheltekqkdelgrlwtvcvqeenlrfmeaetaftqlq 324
QY 324 DIHVPRIKCEVKFASHCLMNHDPDLAKDLTEYLKVRSHDPEEAIRH--DVIVSI----- 374
DB 375 qlhsqsgaelstale-lqnrsqikld---mearnnglqeevqekdqskslnelnlss 379
QY 375 ---VTAAKKOILLVNDHL--LNFVREPLDKRWVRKRAMM---GLAIYIKKY----- 419
DB 380 aasikslqeevsklretiqkiaeaveelvrdqrnalqgeiyclkeelsqigkkgqsmveqv 439
QY 420 ---ALQSAAGDAKQIAWIKDKLLHIYYQNSIDRDLVERIFAQYMWPHNLETTERM-- 474
DB 440 elvglhpesfgsvkelqeenskikfeiresiektaalek-----lemmeklvq 489
QY 475 KCLYLYVATLDLNA-----VRALNEMMKQCNLLR-----HOVKDLDLIKQPKTDAS 521
DB 490 knlllensisdnaeletirgklktlee--asmlaeeksglhsekdmnlslrlqsatens 547
QY 522 VKAIFSKYVWITRNLPDGPKAQDFMKKFTQVLED-----DEK-----I 559
DB 548 -kkiseenmvliensifnanveeleelkslksleeschllnddkttltseresllshidtm 606
QY 560 RKOLEVLYSPCSCCK-----QAEQCVREITKLG---NPKQPTNPFLEMIKFLLEERI 608
DB 607 rkrtedlekeahelkvkvelateresslqki-eelgvslnakd--ceyasfvqfseerm 663
QY 609 ---APVHIDTESISALIKVQNSIDGTVADDEGVPDQAIRAGLELLKLSLFTFHPISF 664
DB 664 ngmestihlhdqncqvreyqveldrahdahielivlqkclqdlw----- 709
QY 665 HSAETTESLACLAKMDDEKVAEALQIFKNTGSKTEEDFPHIRSALLPVHLHHSKKGP 724
DB 710 ---ekssslia-----enqdikeas-kkleklvseleee-----nigkqv 745
QY 725 QAKYAIHCIIHAIFSKETQQAIFELP-----HKSLDPSN-----LEHLITPL 767
```

```
DB 746 qidssincikil----rtgiyqvlmklelipigsgdensrdqrmhmdilnrledmqmtl 801
QY 768 VTI-----GHIAL--LAPDQFAAPKWSVATFIVKDLLMNDRLP----- 804
DB 802 isirdenqhsaienlvlieflrqlkseavgietekhiieeescqqlsfdsrtqkli 861
QY 805 ---GKTTTKLWVPDEEVSPETMVKIOAKIMVMRWLLGMKNHNSKSGTSTLRLLLTILHSD 861
DB 862 fvngelettckv---nggvnrkvlmve-fedfhrqvlqirdy-----tilsgd 905
QY 862 GD--LTEOGKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLEQYOLCALAINDECYQ-- 917
DB 906 nktidekayitk-----stlgleeekckle-----ddisllsetiyqsn 946.
QY 918 ---VRQVFAQKLHKLSRLPLLEYMAICALCAKDPKERRAHRACQLKVKKNINVRREYL 973
DB 947 lliilledvilekl-sgamklnedlrlsivkckleeevrelgdkksadianfqqlvle 1005
QY 974 KOHAASVEKLLSLUPEYVYVYTIHLLAHDPPYKVDIEQLKDVKECLWFVLEILMAKNE 1033
DB 1006 kсна-----ellisarsa-----nvh-leheianvkq-----kekelleamlimisqne 1049
QY 1034 NNSHAFIRKMWENI-----KQTK-----DAQGPDDAKMKNEKLYTVCDVA 1072
DB 1050 kse---Iskaveglecrykeakealedrkqvlrlrgdydeqvkknshsneanlkleadl 1106
QY 1073 MNIIMSKSTTYSLESPK--DPVLPARFFTPQDKNFSNTKNYLP--EMKSFFTPGPKPTTN 1129
DB 1107 mnlm-----eleeikvekenlqelfte-----rneielwesqaatlfgelqisa 1152
QY 1130 V-----LGAVNKPLSAGKOSQTKSS-----RMTVSNASSSN-----PSSPG 1168
DB 1153 vhetillegitnelveacknlesrstkdreieqlkgrvnnledankgndlmckyqaalf 1212
QY 1169 RIKGLDSE---MDHSENEYTMSPLPG---KSKDKRDSOLVSRSELEKPRGRKKT 1220
DB 1213 llikesiglekhamehfeengpattngsfvgisyqetslvsdngf-lelqelhlrika 1271
QY 1221 PVTEOEERLGLMDDLTKLVQEQPKGQSRRKRGHTFASDEQQWPEEKRL----- 1270
DB 1272 leeatlkkllameel-----ktssarrstrngslrkqnheiyseetemitckdivldqv 1324
QY 1271 -----KEDILENEDEQNSPPKKGRRPPKPLGGGTPEEPTMTKTSKGSKKSGP 1321
DB 1325 sdcsygiistrdilkiedhs---leaksqnpkp---gkslseeelvvdkleisdrftd 1377
QY 1322 PAPPEEEEEQSGNTEOKSKSKOHRVSRRAQOARSPESSAIESTOSTPOKGRGRPSKT 1381
DB 1378 pnkdankrkvierlnsdliqlslnhvavedikikve-----teekdekgeneyet 1428

RESULT 15
AAW02258
ID AAW02258 standard; Protein; 1411 AA.
XX
AC AAW02258;
XX
DT 09-MAR-1997 (first entry)
XX
DE Nucleolar/endosomal auto-antigen p162.
XX
KW Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN DE19515514-C1.
XX
PD 12-SEP-1996.
XX
PF 27-APR-1995; 95DE-1015514.
XX
PR 27-APR-1995; 95DE-1015514.
```



PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.



Db 1451 erleikdrdyfelyvtantnlvegkletplqadheedsidrrseemeikvlgeklernqy 1510  
Qy 260 -----LELYN-----IDSHLLLSVL--POLEFKLKS----- 283  
Db 1511 llerlqeeklelsnklellqkemetsvllkddlqklesliseniilkenidttlkhhssd 1570  
Qy 284 -----NONEERLQVVKLLAKMFGAKSELASQNK-----PLWQCVYLGRENDIH----- 326  
Db 1571 tqaqlqktqgelqlaknla--iaasdnpcitqeketsadcvhpieekillilteelhqtkn 1628  
Qy 327 -----VPIRELCVKFASCHLMNHPDLADLTLYLKVRSHPDEE---ATRHD 369  
Db 1629 egekllehekneleaqvelkce--veh-Imksmiesksleslqnekhdtteqqlalakkq 1685  
Qy 370 VIVSIVTAAKKDIILLVNDHL-----LNF---VRETLQKRWVRKEAMMGL 412  
Db 1686 --mqvvtqekkelqgthelthaevdhikeniellgnfkneaqgkttkeqcllnenkeleq 1743  
Qy 413 AQIYKKYALQSAAGKADAKQIAWIKDK-----LHIIYQNSIDRLLVERIFAQYVMPHN 467  
Db 1744 sq-----hrlqc-----eieelmkslkksaletikeseqkvlinqenemvmlmeeikn 1795  
Qy 468 LETT---ERMKCLYLYATLDLNAVKALNEMWKQNLRLHQVKDLDLLIKQ----- 515  
Db 1796 sqrtviaerdqlqdlresvem-sietqddlrkaqeaqqgdkvqelqtsvlqekis 1854  
Qy 516 -----PKTDASVKAIFSKVMVITRNLPDPGKA-----QDFMK 547  
Db 1855 llenqmylnvatvketiserddlnqskhlfsietislsikefaleqaeakdaaar 1914  
Qy 548 KFTQVLEDEKIRKQL-----EVLVSPTCSCKQAECC-----VREITKK--- 586  
Db 1915 ktiditekisnieeqllqgatknetlyeresli-----qckeqialntehlretlkskdl 1970  
Qy 587 -LGPNKQPTNPFLEKIFELLERAPVHDITESIALIKQVKNKSIDGTADDEGVPTDQA 645  
Db 1971 alghmeqerdeaanvial-----tekmslsleeqinenvttikegege----- 2033  
Qy 646 IRAGLELLKVLFSFTHPSFSAETTESILACLKDDKVAEAAALQIFKNTG----- 696  
Db 2014 -----ketfylrpskqssqmeelreslktkdqlgeaeakeiseatneiknlaki 2066  
Qy 697 SKIEDPFHRSAL-LPVLHHSKKGKPPRQAKYAIHCHIAIFSSKETQFAIFPLHKS 755  
Db 2067 slleeeilqnasilneavserenlrhskqglvseleqlsltlksrdhafaq-----skr 2120  
Qy 756 DPSNLEHLITPLVTIGHTALLAPDQFAAPKSWATFTVKDLLMNDRLPGKKT--TKLWV 813  
Db 2121 ekdeavnkiaslaeelkiltkemdefrdskes---lqeqsghiseelctyktelqmlkq 2176  
Qy 814 PDEEVSPTMYKIQAIKMMVWRLGLMKNN-----HSKSGTSTRLTLTLHSDGDLTEQK 869  
Db 2177 qkedinnklaevkvedellqhlslleqldqigmeirneklr-----nyelcekmd 2228  
Qy 870 ISKPDMSRLRLAAGSAIVKLAQEPCEYHIEITLQYQICALAINDECYQVRO-----VFAQK 925  
Db 2229 imekeisavlrl-----mqnepqgeeddvaermidile-srnqeiqlmekisavvseq 2279  
Qy 926 LHKGLSLRLPLEYMAICALCAKDPVERRAHARQCLVKVNNVREYLK----- 974  
Db 2280 -htllslslelq-----keteah-khcm---lnikeslssstlsrfsqslqt 2321  
Qy 975 QHAAYSEKLLPEYVVPYTIHLHADPDYKVKQDIEQLDKVRECLWFVLEILMAKNEN 1034  
Db 2322 ehvkintqlqtlknkfkvvrytaavkedhslik--dye-kdlaa-----eq 2364  
Qy 1035 NSHAFIRKWNENIKOTQDAQCPDDAK-----MNEKIYTVCDVAMNIIIMSKSTYS 1084  
Db 2365 krhdelrlqlcleq-hgrkwsdsaseelkfcetefinelifk-----kanliqsvqddfs 2419  
Qy 1085 -----LESPPK-----DPVLPAREFTOPDKNFSNTKNVLP 1114

Db 2420 evqvflnqvgtlgeelehkkqfmqwlfeefgdllhvdakklsegmqenrriastiqlitk 2479  
Qy 1115 EMKGFF-TPGKPKTTNVLGAVNKPLSAGKQSQTKSSRMETVSNASS---SSNPFSSPGRI 1170  
Db 2480 rlkavvqskiqreiltvynlqfeaklqekkeqnkelmrmehhgppsasvmeenearlilgil 2539  
Qy 1171 KGRLDSSMDHS-----ENE-----DYTMS-----SPLPGKKS 1199  
Db 2540 ktvqdesakklqsrklmlenelnlvkddamhgekvaillqdkllsrnaeeelnamqvkitk 2599  
Qy 1200 KRDDSDLVRSLE---KPRGKKTPTVTEOEELKGMDDLTCLV--OEQPKGSGORSRKRGH 1254  
Db 2600 kqdnlaqamkeienlqkmavkavpykeeiidnl-----ktkvvkiemekikyskatdga 2655  
Qy 1255 TASEDEQOWPEEKLKEDI-----LENEDEQNSPPKGRGRPPKPLGGGTPEKEPTMTK 1310  
Db 2656 ylkscledkeegrlrlkeelrraqadndttvcvpkdyqkastfvvtcggsgivqgstaml 2715  
Qy 1311 SKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHRVSRRAQOARAESPESSATESTQS- 1369  
Db 2716 vlq-----sekaalerelsh---ykkkyhhlsrtmsssedrktktaksdahss 2760  
Qy 1370 -TPQKGRGRPSKTPS---PSQPKN 1390  
Db 2761 htgsshrsgphkcteyrhgvpvtpers 2786  
RESULT 14  
AAG39233  
ID AAG39233 standard; Protein; 1562 AA.  
XX AAG39233;  
XX XX  
DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48514.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX XX  
PD 06-SEP-2000.  
XX XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.



625 glhpesgssvkelqeenksklkeireresiektalie-----lemmeklvqknl 674  
478 YLYATLDLNA-----VKALNEMWKCONLLR-----HQVKDLDDLKQPKTDASVKA 524  
675 llensisdlnaetirgikliee--asmslaeeksglhsekdmllrlsatsens-kk 731  
525 IFSKVMVITRNLPDGPQAQDFMKKFTQVLED-----DEK-----IRKQ 562  
732 lseemvnienslfnanvealeeksklleeschnldkttltseresllshidtmkr 791  
563 LEVLVPTCSCK-----QAEGCVREITKLG---NPKQPTNPFLEMIKFLLERI--- 608  
792 ledlekehaelkvkvelateresslqki-eelgvslnakd--ceyasfvqfsermgm 848  
609 -APVHIDFESTSALLKQVNSIDGTADDEGVPDQAIRAGLELLKVLSTFHPISFUSA 667  
849 estihlhdengcrvreyqveldranhahlelivlqkclqdl----- 891  
668 ETFESLLACLKMDDEKVAEAAQIFKNTGSKTEEDFPHRSALLPVLHHSKKGPPROAK 727  
892 eksssia-----engdikeas-klliekivseleee-----nigkvqvid 930  
728 YAIHCIAFFSKETQFAQIFPL-----HKSLDPSN-----LEHLITPLVTI 770  
931 ssincikil-----rtgiyvlmklelpgigsgdensrdgrnmhdilnrledmqtmllisi 986  
771 ----GHIAL--LAPQFAAPKWSWATFIVKDLLMNDRLP----- 804  
987 rdenqhsalenlileflrqlkseavgtekkileeelesqqlsfsrdetqklifvn 1046  
805 GKKTQKLVWPDEVPETWVKIQAIKMMVRLGLGMNHNHSGSTSLRLTLTLILHSDGQ- 863  
1047 gelttkv---nggvnrekvlmve-iedfhrqlgirdy-----tilggdnk 1090  
864 -ITEGKISKPDMSRLRLAAGSAIVKLAQPCYHEITLEQVQLCALAINDECYQ----- 917  
1091 tidekayltk-----stlgleeekckle-----ddislllsetiyqsnlli 1131  
918 -VROVFAQKLHLKGLSRLRLPLEYMAICALCAKDPKERRAHARQCLVKNINVRREYLKQH 976  
1132 lledvileklsigamklnedldrlsvkckleeevregdklsadlanfqlqvleksen 1190  
977 AAVSEKLLSLPEYVYPTIHLAHDPPDVKVQDIEQLKDVKECLWFVLEILMAENNS 1036  
1191 a----ellsarsa-----nvh-leheianvkq-----kekelleamlmsimneke 1234  
1037 HAFIRKMWENI-----KQTK-----DAQGPDADAKMNEKLYTVCDAVMI 1075  
1235 ---lskaveglecrykeakeaieedrdkqvlrlrgdydeqvkknshsneanlkleadlml 1291  
1076 IMSKSTYSLESFK--DPVLPARFTQDPKNSNTKYNLPP-EMKSFTPGPKPTNV-- 1130  
1292 lm-----eleekvkekenlndelife-----rnelelwesqsatlfgelqisavhe 1337  
1131 --LGAVNKLPLSAGKOSQTKSS-----RMETVSNASSSN-----PSSPGRIK 1171  
1338 tllegtlneivacnleersitlkareqlkgrvnniedankgndlmckyqaifilk 1397  
1172 GRLDSE-----MDHSENEYDTMSSPLPG-----KSKDRDSDSLVRSELEKPRGRKTPVT 1223  
1398 esiqslekhamihefengpattngsfvgsyqetasldvnsdgl-leiqelhirikaatee 1456  
1224 EQBEKLGMDDLTKLVQEQPKSQSRKRGHTASTESDEQWPEERKL----- 1270  
1457 aitkkameel-----ktsaarsrrngslrkqnheiyseeteemitkdilvdqvsdc 1509  
1271 -----KEDILENEONSPPKKGRGRPPKPLGGTTPKEPTMTKSKGSKKSGPPAP 1324  
1510 ssygistrdilkiehdhs---leaksqppk-----gkslseeslvvdkleisdrftdpnk 1562  
1325 EEEEEERQSGNTEQKSKQHRVRRRAQRAESPSSAISTQSTPQKGRGRPSKT 1381  
1563 dankkvlerlnsdliqklsnlhvavedlikve-----teekdekqkeneyet 1610

## RESULT 13

AA01632 standard; Protein; 2954 AA.

XX AC AA01632;

DT 22-JUN-1999 (first entry)

XX Amino acid sequence of centromere-associated protein-E (CENP-E).

XX CENP-E; centromere-associated protein-E; ATPase activity;  
KW plus end-directed microtubule motor activity; chromosome congression;  
KW microtubule binding activity; chromosome movement; mitosis;  
KW cell proliferation; tumor; metastasis; vascular malfunction;  
KW inflammatory disease; immune disease; angiogenesis; hypertension;  
KW restenosis; fungal infection; selective herbicide; fungicide;  
KW insecticide; plant growth regulator; activator; cancer cell marker.

XX Xenopus sp.

XX WO9913061-A1.

XX PD 18-MAR-1999.

XX PF 10-SEP-1998; 98WO-US19231.

XX PR 11-SEP-1997; 97US-0058645.

XX (REGC ) UNIV CALIFORNIA.

XX PI Cleveland DW, Goldstein LSB, Sakowicz R, Wood KW;

XX WPI; 1999-229233/19.

XX DR N-PSDB; AAX26819.

XX Centromere-associated protein-E and related nucleic acid

XX Claim 5; Page 66-67; 77pp; English.

XX The present sequence represents CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, ATPase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic, bioagricultural and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; vascular malfunction; inflammatory and immune diseases; angiogenesis; hypertension; restenosis; and fungal infections); also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells.

XX Sequence 2954 AA;

Query Match 2.7%; Score 192; DB 20; Length 2954;

Best Local Similarity 18.5%; Pred. No. 0.0011;

Matches 294; Conservative 239; Mismatches 575; Indels 478; Gaps 69;

QY 110 ITRQLKGLSDT---KSPQFNRYFYLLLENIAWVKSYNICFELEDSEIFITQLYRTLFSVIN 166

Db 1374 ickkingleleikgseeevlksmlenlk-----ednnklkeqa-----eys 1417

QY 167 NGHNQKVHMHVMDLMSSTICEGDTVSQELLDVNLVPAHKNLNKQAVDLAKALKRPA 226

Db 1418 skeng---fsleevfsg-----sqklvdei-----evikaqikaee 1450

QY 227 QAIEPVIITFF-----NQVLMGLGTSISDLSEHFDLI----- 259







897 ssincikil-----rtgiyvImkleiIpiqsgdensrdqrmhdiIlnrlmdmqtmllsi 952  
771 ---GHIAL--LAPDOEAPKSWVATFIVKDLLMNDRLP----- 804  
953 rdenqhsaenlvleflrlkseavgietekkleeelesccqqlsfrdetqklfvn 1012  
805 GKTKTKLWPDDEEYSPETMVAIOATKMMVRWLLGMMNHKSSTGLRLLTILHSDGD- 863  
1013 geltkv---nggvnrekvlmve-ledfhrqvlrlday-----tlqgdunk 1056  
864 -LTEGKTSKPDMSRLRLAAGSAIVKLAQEPYHEIITLEQYQCALAINDECYQ----- 917  
1057 tldekayitk-----stlqleeeckle-----ddislllsetiyqsnlii 1097  
918 -VRQVFAQKLHKLRLPLPLEYMAICALCAKDPKERRAHARQCLVKNINVRREYLKQH 976  
1098 lledvilckl-sgamklnedrlsivckleeevregldkkaadianfqlqvvlksn 1156  
977 AAVSEKLLSLLPEYVVPVTIHLAHPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNS 1036  
1157 a-----ellisarsa-----nvh-leheianvkvq-----kekelleamlmisinqkse 1200  
1037 HAFIRKMVENI-----KQTK-----DAQGPDDAKMNKLYTVCDAVAMI 1075  
1201 ---lskaveglecrykeakaieedrkvrlrgdydeqvkknshsneanikleadlmnl 1257  
1076 IMSKSTYVLSLSPK--DPVLPARFTQDKNFSNTKNVLP-EMKSFTTPGPKTKTNV-- 1130  
1258 lm-----eleekvekenlnqelfte-----rneielwesqsatlfgelgisavhe 1303  
1131 --LGAVNKPLSSAGKOSOTKSS-----RMETVSNASSSSN-----PSSPGRIK 1171  
1304 tlleqtlneIveacknlsrsltkdreqlkgrvnniedankgndlmckyayaqafilk 1363  
1172 GRLDSE-----MDHSENEYTYMSSPLPG-----KSKDRDDSDLVSRSELEKPRGRKKTPTV 1223  
1364 esiglekhamlhefengpattngsfvgsygetaslvdsngdf-leiqelhlrikaee 1422  
1224 EOEELGHDDTLKLVQEQPKGSQSRKRGHTASDESQOWPEEKRL----- 1270  
1423 altkklameel-----ktsarrsrringslrkqnheiyseetemitkdivldqvadc 1475  
1271 -----KEDILENEQNSPPKKGRRPPKPLGGGTPEKEPTMTKSKKSKKSGPPAP 1324  
1476 ssyigstrdilkiedhs---leaksnppk---gkslseeslvvdkleisdrftdpnk 1528  
1325 EEEEEERQSGNTEQKSKOHRVSRRAQRAESPESAIESTQSTQKGRGRPSKT 1381  
1529 dankrvlierlnsdIqklsnhvavedikikve-----teekdekkeneyet 1576

RESULT 12  
AAG36712  
ID AAG36712 standard; Protein: 1744 AA.  
AC  
XX AAG36712;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45030.  
DE  
XX  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX

Arabidopsis thaliana protein fragment SEQ ID NO: 45030.

Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.



```
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
10-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.
07-SEP-1999; 99US-0152363.
10-SEP-1999; 99US-0153070.
13-SEP-1999; 99US-0153758.
15-SEP-1999; 99US-0154018.
16-SEP-1999; 99US-0154039.
20-SEP-1999; 99US-0154779.
22-SEP-1999; 99US-0155139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
05-OCT-1999; 99US-0157753.
06-OCT-1999; 99US-0157865.
07-OCT-1999; 99US-0158029.
08-OCT-1999; 99US-0158232.
12-OCT-1999; 99US-0158369.
13-OCT-1999; 99US-0159293.
13-OCT-1999; 99US-0159294.
13-OCT-1999; 99US-0159295.
14-OCT-1999; 99US-0159329.
14-OCT-1999; 99US-0159330.
14-OCT-1999; 99US-0159331.
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.78; Score 192; DB 21; Length 1710;
Best Local Similarity 18.98; Pred. No. 0.00051;
Matches 305; Conservative 260; Mismatches 536; Indels 456; Gaps 76;

Qy 20 KEISDKI-SKEEMVRRUKMVYKTFMDMDQDSEEEKELYL-----NLALHLASDFDLKHPG 73
Db 161 keinakviseseraskaeaeivalkdalskvgaekaslaqfdqpnleklsnlesevsraq 220
Qy 74 KDVRLVACCLADIFRIYAPAPYTPDKLADIFMTITRQLKGLDTPKSPQFNRYFYILLE 133
Db 221 edsrville-----ratraaeavet-----lreslskvevekesllgyqqclq 263
Qy 134 NIAWKSYNICFELEDSNEIFTOLYRTLFSVINNGHOKVHMHVMDLMSIICEGDTVS- 192
Db 264 nla-----dledrislaqkeagevderanraeaeatlakqslvssetdkeaalvqy 314
Qy 193 QELLDTVLVNLPVPAHKNLQAYDLAKALKRTAQAIPIYITTFNQVLMGLKTSISDLS 252
Db 315 qqclkti-snl---eerlhkaeed--srlnqraenaagevesl-----kqkvskli 360
Qy 253 EHVFDLILELYNIDSHLLSVLPQLEFKL-KSNONEERLQ-----VVKL----- 295
Db 361 ee-----neayelgyqqcltdladlkklfhaeetqrlsreiedgvaklkfaeekcvvl 415
Qy 296 -----LAKMEGAKDSELASONKP---LWQCY-----LGRFNDDIH 326
Db 416 ersnqnhseldgllekignqshelteqkelgrlwtcvqeenlrmeaetaftqlqlh 475
Qy 327 VPIRLECVKFASHCIMNHPDLAKDLTEYLKVRSHDPEAIRH--DVIVSI----- 374
Db 476 sqsqeelstlae-lqnrsqilkd---mearnnglqevqeqakqskslnelnissaas 530
Qy 375 VTAAKKDILLVNDHL--LNFVRETRTLQKRWVRKEAMM---GLAQIYKY----- 419
Db 531 ikslqeevsklretiqkleaeevlrvdqrdnalqleqylckeelsqigkghdmsveqvelv 590
Qy 420 ALQSAAGKDAAKQIAWIKDKLLHIYQNSIDRLIVERIFAQYVMPHNLLETTERM--KCL 477
Db 591 glhesfgsvkqeensklkeireslektaliek-----lemmeklvqknl 640
Qy 478 YLYATLDLNA-----VKALNEMMKQNLRLR-----HQVKDLDDLKOPKTDASYKA 524
Db 641 llensisdlnaeletirgkltlee--asmslaeeksglhsekdmllsrqlsatens-kk 697
Qy 525 IFSKVMVITRNLPDPGKAQDFMKFTQVLED-----DEK-----IRKQ 562
Db 698 lseenmvlenlsfianveleeksklsleeschllndkttltseresllshidtmrkr 757
Qy 563 LEVLVSPTCSCK-----OAEQCVREITTKLG--NPKOPTNPFLEMIRKFLLERI--- 608
Db 758 iedlekehaelkvkvelateressiqki-eelgvslnakd--ceyasfvqfseermngm 814
Qy 609 -APVHIDTESIALIKQVNSIDGTADDEGVPTDQAIRAGLELLKVLFSLTFHPSFHA 667
Db 815 estihhlqdenqcrvreyqveldrahdahielvqlcklqdlw----- 857
Qy 668 ETFESLLACLKMDDEKVAEAAALQIFKNFGKIEEDFPHIRSALLPVLLHHKSKKPPROAK 727
Db 858 ekssslia-----enqdikeas-killeklvsleee-----nigkqvqld 896
Qy 728 YAIHCIIHAIFSSKETOFQAFIEPL-----HKSIDPSN-----LEHLITPLVTI 770
```



1194 ---lskaveglecrykaeiaeedrdkgvrlrlrgdydeqvkknshsneanlkleadlmm1 1250  
99US-0134219.  
PR 14-MAY-1999;  
14-MAY-1999;  
99US-0134221.  
PR 14-MAY-1999;  
14-MAY-1999;  
99US-0134370.  
PR 18-MAY-1999;  
18-MAY-1999;  
99US-0134768.  
PR 19-MAY-1999;  
19-MAY-1999;  
99US-0134941.  
PR 20-MAY-1999;  
20-MAY-1999;  
99US-0135124.  
PR 21-MAY-1999;  
21-MAY-1999;  
99US-0135353.  
PR 24-MAY-1999;  
24-MAY-1999;  
99US-0135629.  
PR 25-MAY-1999;  
25-MAY-1999;  
99US-0136021.  
PR 27-MAY-1999;  
27-MAY-1999;  
99US-0136392.  
PR 28-MAY-1999;  
28-MAY-1999;  
99US-0136782.  
PR 01-JUN-1999;  
01-JUN-1999;  
99US-0137222.  
PR 03-JUN-1999;  
03-JUN-1999;  
99US-0137528.  
PR 04-JUN-1999;  
04-JUN-1999;  
99US-0137502.  
PR 07-JUN-1999;  
07-JUN-1999;  
99US-0137724.  
PR 08-JUN-1999;  
08-JUN-1999;  
99US-0138094.  
PR 10-JUN-1999;  
10-JUN-1999;  
99US-0138540.  
PR 14-JUN-1999;  
14-JUN-1999;  
99US-0138847.  
PR 16-JUN-1999;  
16-JUN-1999;  
99US-0139119.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139452.  
PR 16-JUN-1999;  
16-JUN-1999;  
99US-0139453.  
PR 17-JUN-1999;  
17-JUN-1999;  
99US-0139492.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139454.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139455.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139456.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139462.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139463.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139750.  
PR 18-JUN-1999;  
18-JUN-1999;  
99US-0139763.  
PR 21-JUN-1999;  
21-JUN-1999;  
99US-0139817.  
PR 22-JUN-1999;  
22-JUN-1999;  
99US-0139899.  
PR 23-JUN-1999;  
23-JUN-1999;  
99US-0140353.  
PR 23-JUN-1999;  
23-JUN-1999;  
99US-0140354.  
PR 24-JUN-1999;  
24-JUN-1999;  
99US-0140695.  
PR 28-JUN-1999;  
28-JUN-1999;  
99US-0140823.  
PR 29-JUN-1999;  
29-JUN-1999;  
99US-0140991.  
PR 30-JUN-1999;  
30-JUN-1999;  
99US-0141287.  
PR 01-JUL-1999;  
01-JUL-1999;  
99US-0141842.  
PR 01-JUL-1999;  
01-JUL-1999;  
99US-0142154.  
PR 02-JUL-1999;  
02-JUL-1999;  
99US-0142055.  
PR 06-JUL-1999;  
06-JUL-1999;  
99US-0142390.  
PR 08-JUL-1999;  
08-JUL-1999;  
99US-0142803.  
PR 09-JUL-1999;  
09-JUL-1999;  
99US-0142920.  
PR 12-JUL-1999;  
12-JUL-1999;  
99US-0142977.  
PR 13-JUL-1999;  
13-JUL-1999;  
99US-0143542.  
PR 14-JUL-1999;  
14-JUL-1999;  
99US-0143624.  
PR 15-JUL-1999;  
15-JUL-1999;  
99US-0144005.  
PR 16-JUL-1999;  
16-JUL-1999;  
99US-0144085.  
PR 16-JUL-1999;  
16-JUL-1999;  
99US-0144086.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144325.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144331.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144332.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144333.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144334.  
PR 19-JUL-1999;  
19-JUL-1999;  
99US-0144335.  
PR 20-JUL-1999;  
20-JUL-1999;  
99US-0144352.  
PR 20-JUL-1999;  
20-JUL-1999;  
99US-0144632.  
PR 20-JUL-1999;  
20-JUL-1999;  
99US-0144884.  
PR 21-JUL-1999;  
21-JUL-1999;  
99US-0144814.  
PR 21-JUL-1999;  
21-JUL-1999;  
99US-0145086.  
PR 21-JUL-1999;  
21-JUL-1999;  
99US-0145088.  
PR 22-JUL-1999;  
22-JUL-1999;  
99US-0145085.  
PR 22-JUL-1999;  
22-JUL-1999;  
99US-0145087.  
PR 22-JUL-1999;  
22-JUL-1999;  
99US-0145089.  
PR 22-JUL-1999;  
22-JUL-1999;  
99US-0145192.  
PR 23-JUL-1999;  
23-JUL-1999;  
99US-0145145.  
PR 23-JUL-1999;  
23-JUL-1999;  
99US-0145218.  
PR 23-JUL-1999;  
23-JUL-1999;  
99US-0145224.

1194 ---lskaveglecrykaeiaeedrdkgvrlrlrgdydeqvkknshsneanlkleadlmm1 1250  
1076 IMSKSTTYSLESPK--DPVLPARFFTOPDKNFNTKNVLP--EMKSFFTPGPKPTNV-- 1130  
1251 Im-----eleeikvekenIngelite-----rnelwesdaatlfgelqisavhe 1296  
1131 --LGAVNKPSSAGKQSQTKSS-----RMETVSNASSSSN-----PSSPGRIK 1171  
1297 tlllegtlneveacknlestrtkdreieqIkgvnniedankgndlmckyayaqailk 1356  
1172 GRLDSE-----MDHSENYDMSSPLPG-----KSDKKRDDSLVRSELEKPRGRKPTVT 1223  
1357 esiqslekhamIhefengpattngsfvqysetsaslvdsngf-leiqelhlrikaatee 1415  
1224 EQEELGMDLTKLVQEQPKGSQRKRGHWTASESDQOWPEEKRL----- 1270  
1416 aitkkIameel-----ktssarrsrrngslrkqnheiyseetemitkdilvdqvsdc 1468  
1271 -----KEDILENEDEQNSPPKKKGRPPKPLGGTPEEPTMTKSKGKKKSGPPAP 1324  
1469 ssygistrdilkiehdhs---leaksqppk---gkslseeslvvdkleisdrftdpnk 1521  
1325 EEEEEERQSGNTEQKSKQHVRERRAQRAESPESSAIESTQSTPQKGRGRPSKT 1381  
1522 dankrkvlernsdIqklsnlhvavedIkikve-----teekdekgkeneyet 1569

RESULT 11  
AAG36713  
ID AAG36713 standard; Protein; 1710 AA.  
AC AAG36713;  
AA AAG36713;  
DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45031.  
DE Arabidopsis thaliana  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS EP1033405-A2.  
PN EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.







2426 sgtilrepttksvpvnlpersptdspre 2454  
RESULT 10  
AAG36714  
ID AAG36714 standard; Protein; 1703 AA.  
XX  
AC AAG36714;  
XX  
DI 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45032.  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45032.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 04-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 06-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147933.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.



24-OCT-1994; 94US-0328254.  
22-OCT-1993; 93US-0141239.  
(TEXA ) UNIV TEXAS SYSTEM.  
Lee W, Zhu X;  
WPI; 1998-109817/10.  
N-PSDB; AAV09076.  
New isolated mitosis protein and gene - useful for, e.g. developing products for therapy and diagnosis of hyper-proliferative disorders such as cancers or psoriasis  
Claim 1; Column 40-52; 43pp; English.  
This is the amino acid sequence for mitosis, a phosphoprotein necessary for the cell to enter mitosis. The protein's degradation is necessary for the cell to advance into the next stages of mitosis. The mitosis protein, can be used to control the growth of cells. An anti-mitosis antibody, a mutant or a non-functional analogue of mitosis can inhibit the mitotic cell cycle by preventing the cells from entering the M phase, and over expression of mitosis or its functional equivalent, would inhibit the cycle by preventing cells from leaving the M phase. Antagonists to this protein can be used to control hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer and various leukemias and lymphomas). Reintroduction or supplementation of lost mitosis function by introduction of the protein or nucleic acid encoding the protein into a cell can restore defective chromosome segregation, which is a marker of progressing malignancy. Malignant proliferation of cells can then be halted. The protein can also be used for the detection and diagnosis of hyperproliferative cells.

Sequence 2482 AA:  
Query Match 2.7%; Score 194.5; DB 19; Length 2482;  
Best Local Similarity 19.4%; Pred. No. 0.00058;  
Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;  
19 VKETSDKI-----SKEEMVRLKRVVTFMDMDODSEEEKLY-----56  
1172 mkeldsklhieqvmkcieacielekvigeklenskdslekleyfscdqhellqrvtvs 1231  
57 --LNLALHLSDFLKHGKDVRLVACCLADIFRIYAPAPYTPSPDKLXDFMFIPTROL 114  
1232 eglnsdlembad---ksredidgnva-----kvndswkerfidvenel 1272  
115 KGLDRTKSPQFNRYFLLLENIAWVKSYNICFELEDSN--EIFTQLYRTLFSVINNGHNQK 172  
1273 srirsekastehealyleadlevvqtcklekdnenkqvicleeel-svtsernql 1331  
173 VHMHMVDMLSIIEGTVSQELLDVTLVNLVPAHKNLNQAYDKALKALRTQAQIEPY 232  
1332 --rgeldtmskttaldqlsekemek--tgeleshqseclhciaevaevekeltellq-- 1385  
233 ITTFFNQVLMIGKTSISDLSFHVFDLI-----LELYNTDLSHLLSVLPOLFELKLSNDNE 287  
1386 -----Cissdseilkkdkhlqeklslekdqsaltkcelengiaqlnke 1432  
288 ERLQV--VKLIAMFGARDSLASONKPLMOCYL--GRFNDIHVPIRLECVKFASHCLMN 343  
1433 kelliakesesiqarlsesdyeklnvskalealvekegef-----alrlsstgeevhqlrr 1487  
344 HPDLAKDLTEYKVRSHDPEAIRHDVIVSVTAARKDILLVNDHLLNFRVREPLDK---400  
1488 g-----ieklrvrieadekqh--iaeklkererendslkdkvenlerelqmseenq 1538  
401 -----RWRVRKEAMMGLAQIYKRYALQ-----SAAGRDAQAQIAWIKDKLLHIY 444

Db 1539 elvildaenskaevetlktqieemarslkvfeldvltrsekenltkqigekqgqlseld 1598  
QY 445 YQNSIDDRLLVERIFAQYMPVPHNLETTERRMKCLYLYATLDLNAVAKALNEMWKQNLRLH 504  
Db 1599 klissfksllcekeqaeiqikeestavem-----lqn 1631  
QY 505 QVKDLLDLIKOPTDASVKAIFSVMVITRNLPPD-----GKAQDFMKFFQVLEDEKI 559  
Db 1632 qikelneavaalcgd-----qeimkatqegldppieehqlrnsieklrleadek- 1683  
QY 560 RKQLEVLVSTPCSCQAEGCVREITTKLGNPKQPTNPFLEMIKELLERAPIVHDITESIS 619  
Db 1684 -kgicvliq---qikesehadlikgrvenlere-----leartnqehaa---leaeensk 1731  
QY 620 ALIKQVKNKSIDGTADDEDEGVPTDQIRAGLELLKVLSTFHPISFSAETFESILACLAKM 679  
Db 1732 gevetlkakiegmt-----qslr-glel-----dvtirsek--enltneqlk 1771  
QY 680 DDEKVAEAL--QIFKNTGSKIEEDFPHIR---SALLPVLHHSKKGPPROAKYAIHCIIH 734  
Db 1772 egeriseleinsfenilqekeqekvgmekstamemlqtqikeinerva--alhdndq 1829  
QY 735 AIFSSKETQQAIFELPH-----KSLDPSNLEHLITPLVTIGHIALPADQFAAPWKS 787  
Db 1830 eackakeqnissqvecelelekaqlqgldaknnyivlqssvng-----1873  
QY 788 WYATFIVKDLLMNDRLPGKKTTLKLPVDEVPETWYKIAQKMMRWLLCMKNHNS--K 845  
Db 1874 -----liqevdgkq--klekdeels-rlknqiddqevklsgvegehqqlwk 1920  
QY 846 SGTSTLRLTLILHSDGLTEQGIKISKPDMSRLAAGSAIVKLAQBPCYHEITLLEQYQ 905  
Db 1921 eqnlelnltveleqkiqvlqsknaslqg-----tlevlq 1955  
QY 906 LCALAINDECYQV---ROVFAQKLHLGLSR---LRLPLEYMAICALCAKDPV---KERRA 956  
Db 1956 ssyknlenlelctkmdkmsrvekvnkmtaketelqremhemaqktaelqeelsgeknrla 2015  
QY 957 HARQCLVKININVRREYLKQHAHVSEKLLSLLP-----EYVVPYTHLLAHD 1002  
Db 2016 gelqlleeksskdqkeltlenseklksldcmhkdqvekegkvreeiaeyqirl--he 2073  
QY 1003 PD-----YVKQDI--EQLKDVKECL---WVLEILMAXNE--NNSHAFIRK 1042  
Db 2074 aekhqallidtnkqyeiqtyrekitkskeclssqgleidlilkskeelnnslkattq 2133  
QY 1043 MVENIKOTKDAQGDDAKMNEKLYTVCDVA---MNIIMSKSTTYSLSPKDPV-----L 1093  
Db 2134 illeektkt---mdnlkyvynqlkkenraqgkmklli--ksckqleeekeilqelsql 2187  
QY 1094 PARFETQDKNFNSNTK--NYPPEMKSFETPGPKPTNW-----LGAVNKPLSSAGK 1143  
Db 2188 qaakekktgtvmdtkvdeitteleketleektkeadeyldkycsilishekleake 2247  
QY 1144 QSQTK-----SSRMETVSNASSSNSSPNPGRIKRLDSSEMDHSENEEDYTHSSPLP-----1194  
Db 2248 mletqvahlcsqgskdsgspilgvpvg-----pspipsvte 2286  
QY 1195 -----GKSKDRDDSDLVRSLELEKPRGKKTPTVTEQEELGMDLTKLVQEQKPKGSQ 1247  
Db 2287 krissggnksgkrq-----rsggiwengggtpat-----pesfskksk 2326  
QY 1248 RSRKRG--HTASEDEQOOWPEEKRLKEDILENEDQNSPPKKGKRGPPKPLGGGTP--KE 1304  
Db 2327 kavmsghpa-----edtegtefepegilpevvkkgfadiptgktspyllr 2371  
QY 1305 EPTWKT---SKKSGSKSGPPAPPEEEEEERQSCNTQKSKSKOHRVSRRAQQAESPES 1361  
Db 2372 rttmatrtprlaaqklalsplslglenlaesskptaggsrsgkvkvaqr-----spvd 2425  
QY 1362 SAI-----ESTQSTPQKGRGRPSKTPSPSQ 1386







2216 g-----ieklrvieadekkqlh--iaeklkererendsalkdkvenlerelqmseenq 2266  
402 -----WVRKEAMMGLAOIYKVALSAAGKD-AAKOIAWIKOKLLHIY 444  
2267 elvildaeanskaevetlktqleamarskifeldvltrsekentlktqiekgqqlseid 2326  
445 YONSIDRLVVERIFAQYVPHNLETTERMKCLYLYATLDNAVAKALNEMWKCQNLRH 504  
2327 kllsfskllseeqaelikeestavem-----lqn 2359  
505 QVKDLILLIKOPKPTDASVKAIFSKVMVITRNLPD-----GKAQDFMKKFTOVLEDEDEKI 559  
2360 qikelineavaalcgd-----qelmatedslpdpieehqrlnstekrlrarleadek 2411  
560 RKQLEVLVPTSCCKQAGCVREITKLGPNKQPTNPLEMIKFLLEIRIAPVHIDTESIS 619  
2412 -kqlcvq----qkesehahdlkgrvenlere---lelartngqehaa---leaensk 2459  
620 ALIKOVNKSIDGTADDEGVPTDOATRAGLELLKVLSTFHPISFHSAAETFESELLACLKM 679  
2460 geveltkaklegmt-----qslr-giel-----dvvrtirsek--enltneqlq 2499  
680 DDEKVAEAL--QIFKNTGSKIEEDFPHIR---SALLPVLHHKSKGPPROAKYAIHCIIH 734  
2500 egeriseleifinsfennllqekedqvmqkskstamemlqtqikelneriva--alhnhdq 2557  
735 AIFSSKETQFAIPEPLH-----KSLDPNLEHLIPLVTIGHIALLPDQFAAPWKS 787  
2558 eachakeqnisqvecleleakqllqgldeaknyivlqssvgk----- 2601  
788 WATFIVKDLMDRLPGKTTKQLWPDVEVSPETWYKIAIKMVMRWLLGMKNHNS--K 845  
2602 -----llqevdgkq--klekdeels-rlknqigdqqlvsklsqvegehqqlwk 2648  
846 SGTSTLRLTLTILHSDGDLTEQGIKSRPDMRSLRLAAGSAIVKLAQPCYHEIITLEOYQ 905  
2649 eqnlelmitveleqkqlvqsknaslqd-----tlevlq 2683  
906 LCAIAINDECYQV---ROVFAKRLHKLGLSR---LRLPLEYMAICALCAKDPV---KERRA 956  
2684 ssyknlenelctkmdkmsfvekvnmktaketelqremhemaqtaelqeelsqeknrla 2743  
957 HARCLVKNINVRREYLIKQHAANVSEKLLSLP-----EVVVPYTHLHAHD 1002  
2744 gelgllleekskdqkeltlenselskldcmhkdqvekgvreesiaeyqirl--he 2801  
1003 PD-----YVKVDI--EOLKDVKECL--WVLEILMAKNE--NNSHAFIRK 1042  
2802 aekhqallldtnkqyeidqyrekitskeecissqkleidilkskeelnnslkatq 2861  
1043 MVENIKOTKDAQGDDARMNEKLYTVCDA---MNIIMSKSTTYSLESKDPV-----L 1093  
2862 ileelkktk---mdnlkyvnlqkkeneragkmlili--ksckgleeekelqkelsql 2915  
1094 PARFTQDKNFSWK--NYPPEKNKSFPTPGKPTNV-----LGAVNKLPLSSAGK 1143  
2916 qaaqekqktgtvmdtkvdelteikelteektkeadeyldkycslilishlekake 2975  
1144 QSOTK-----SSRMETVSNASSNSPSPGRIKGRLDSEMDHSENEDYTMSSPLP---- 1194  
2976 mletqvhalsqqgkqdsrgslpgvpyg-----pspipsvte 3014  
1195 -----GKKSDDKDDSLVSELEKPRGKKTPTVEQEELKGMDDLTKLVOEQKPGSQ 1247  
3015 krlssgqnkaagkq-----rsggiwengrtpat-----pesfskksk 3054  
1248 RSRKRG--HTASESDQWPEEKRLKEDILENEDEQNSPPKGRGRPPKPLGGGTP--KE 1304  
3055 kavmsghpa-----edtegtefepglpevvvkgfadiptgktspyllr 3099  
1305 EPTMKT---SKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHRVSRRAQRAESPES 1361  
3100 rttmatrsprlaaqaikalsplsigkenlaesskptaggsrskvqvaqr-----spvd 3153

OY 1362 SAI-----ESTQSTQKGRGRPSKTPSPSQ 1386  
Db 3154 sgtlrepttksvpnnlpsrpsdpre 3182

## RESULT 8

AAR72826  
ID AAR72826 standard; Protein; 2482 AA.  
XX AAR72826;  
AC AAR72826;  
XX 27-FEB-1996 (first entry)  
XX Human mitosis.  
DE Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;  
KW inhibition.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 1480..1659  
FT /label= internal\_repeat  
FT Region 1660..1839  
FT /label= internal\_repeat  
XX WO9511309-A2.  
XX  
XX 27-APR-1995.  
XX 24-OCT-1994; 94WO-US12162.  
XX 22-OCT-1993; 93US-0141239.  
XX (TEXA ) UNIV TEXAS SYSTEM.  
XX Lee W, Zhu X;  
XX WPI; 1995-170229/22.  
XX N-PSDB; AAQ86851.  
XX Purified mammalian protein mitosis and agents that bind it and  
PT inhibit its action - used to promote cell growth or to inhibit cell  
PT division and/or proliferation  
XX Claim 4; Fig 8B; 61pp; English.  
XX AAR72829 is human mitosis. Mitosis is involved in the regulation of  
CC the mammalian mitotic cell cycle. Mitosis as with E2F-1 (see AAR72824)  
CC interacts with the retinoblastoma protein (the retinoblastoma tumour  
CC suppressor gene product). Mitosis is first synthesised at the G1/S  
CC boundary, it is then phosphorylated from S through M phase, and during  
CC mitosis, is closely associated with the centromeres/kinetochores at the  
CC mitotic spindle poles. Mitosis is necessary for a eukaryotic cell to  
CC enter the M phase of the mitotic cell cycle and its degradation is  
CC necessary for a cell to advance on to the next stage. Mitosis is thus  
CC useful for controlling cell growth as overexpression of mitosis prevents  
CC a cell from exiting the M phase.  
CC An anti-mitosis antibody, antibody fragment or a phosphorylated mitosis  
CC muten ( or nucleic acid encoding it) can also be used to inhibit cell  
CC division which is particularly useful for the study of the cell cycle.  
CC A further use is to control hyperproliferative cells, and so control  
CC diseases such as psoriasis and breast cancer. It can also be used to  
CC block gametogenesis of an immature gamete.  
XX SQ Sequence 2482 AA;

Query Match 2.7%; Score 194.5; DB 16; Length 2482;  
Best Local Similarity 19.4%; Pred. No. 0.00058;  
Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;



DB	928	k-----	eccklltp-----	ekkllegq-----	ldclcnak	954
QX	803	LPGKTTTKLWVPD--	EEVSPETWAIQIAIKMWVLLGKMNHSHKSGTSLRLTLTILHS	860		
DB	955	teadk--krcvkdpldkqkvlyak-esvkayldcvstrarnekekk-----	997			
QX	861	DGDLTSSQKISPDMSRLRANGSAIVKLAQPCPYEHITITLEQYQOLCALAINDECQVVRQ	930			
DB	998	-----	ecckllltpcakllleekesl--kaydcqlsgarneeraceklltpe--ark	1047		
QX	921	VFAQKLHKGL-----	SRLRPLEYMAICALKADPKVKERRAHARCLVK--NINVRRE	971		
DB	1048	illevkvsikayldcvstrarnekekkceckllltpearkflakqvlncleakagneerka	1107			
QX	972	YLKHOAAVSEKLLSLPEYVVPYTTTHLLAHDPYVKVQDIEQLKDXDKCELWFLEILMAK	1031			
DB	1108	clkn-----	lpklqg--nilak-----eslkaykdcl-----sqar	1137		
QX	1032	NENNSHAF-----	IRKMVENIKOTKDAQDPDAKMNKILYTVCDVAMNIIIMSKSTYS	1084		
DB	1138	neeracekllltpearkille-----	gevkssvayldcvstrarnekekkceck	1186		
QX	1085	LESFPDVLPAEFFQDPKNFNS--TKNYLP-----	PEWKSEFFPGPKPTT	1128		
DB	1187	lltpearkflakelqgkkaikclknadpnadraaimkclgdlsdeckllylgearekav	1246			
QX	1129	NVLGAVNKKPLSSAGKQSQPKSSRMETVSNASSSSNPSPGRIKGRLBDSSE--MDHSEN--E	1185			
DB	1247	adclamaakteekrcqnlysdllceiqnkrtnkqnlksterlhaseclndldpdt	1306			
QX	1186	DYTMSSPLPGKXSDDR-----	DDSLVRSELEKPR-----GRKKTPTVTEQEEKLG	1230		
DB	1307	gealeqcleqdsderallilqrgadevdllylsdlrnrktfdmaakgpyllpmdfkog	1366			
QX	1231	MD-----	DLTKLVQZOKPFGSGORSRKHCHRTASESD--BOOWPEEKRLKEDILENEDE	1280		
DB	1367	gdiatlnatnvdadkiasnpl-----	yasieplakqyetehtklknle----	1412		
QX	1281	QNSPPKKGRRPPKPLGGGTKEBPPTMKTSGKSSKSGPPAPEEPEERQSGNTEQK	1340			
DB	1413	-----	aklaalggn--kkdddkksskstaekaennkidkdvatakniseia	1460		
QX	1341	SKSKQHVRVSRRAQORAE--	SPRESSAISTQSTPOK-----GGRGPS	1379		
DB	1461	lknkkesgfevdengnpiddlkkaekadetspvkqafiqksdpt	1505			

RESULT 7

AA99795	AA99795 standard; Protein; 3248 AA.
LD	
XX	
XX	
AC	
AC	
AA99795;	
XX	
XX	
DT	
DT	08-OCT-1996 (first entry)
XX	
DB	Kinetochores protein CENP-F.
XX	
XX	
KW	Kinetochores protein; CENP-F; cell cycle; cancer; diagnosis;
KW	autoimmune antibody.
XX	
XX	
OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FT	1..200
FT	/label= Extended_coiled_structure
FT	280..1350
FT	/label= Extended_coiled_structure
FT	1380..1610
FT	/label= Globular_domain
FT	/note= "globular domain consists of 2 direct
FT	repeats of 95 amino acids"
FT	1620..1750
FT	
FT	Domain

FT	/label= Extended_coiled_structure
FT	1850..2990
FT	Domain
FT	/label= Extended_coiled_structure
FT	3048..3248
FT	Domain
FT	/label= C-terminal_domain
FT	/note= "the C-terminal domain is predicted to
FT	form a proline-rich (10.6%) highly
FT	basic (pI 10) globular domain"
XX	
XX	WO9617867-A1.
PN	
XX	
XX	13-JUN-1996.
PD	
XX	
PF	08-DEC-1995; 95WO-US16216.
XX	
PR	09-DEC-1994; 94US-0353700.
XX	
XX	(FOXC-) FOX CHASE CANCER CENT.
PA	(UYTE-) UNIV TECHNOLOGIES INT INC.
XX	
PI	Rattner JB, Yen TJ;
XX	
XX	WPI: 1996-287116/29.
DR	N-PSDB; AAT34578.
XX	
PT	DNA encoding kinetochore protein - used as a marker for the G2 and M
PT	phases of a cell cycle, partic. for detection of malignant diseases
XX	
PS	Claim 12; Page 41-54; 72pp; English.
XX	
CC	A 372 kDa human kinetochore protein, CENP-F (AAR99795), is detected
CC	by immunofluorescence microscopy only during the G2 and M phases
CC	of a cell cycle. It is the product of a cDNA clone (AAT34578)
CC	isolated from a breast carcinoma cDNA library. Recombinant CENP-F
CC	can be produced by expression in prokaryotic or eukaryotic host
CC	cells. CENP-F can be used to detect autoimmune antibodies to
CC	the protein, which may provide an early diagnosis for the onset
CC	of various malignant diseases. Use of CENP-F as a cell cycle
CC	marker allows the specific detection of G2 and M phase cells.
XX	
SQ	Sequence 3248 AA;

```
Query Match      2.8%; Score 201.5; DB 17; Length 3248;
Best Local Similarity 19.5%; Pred. No. 0.00028;
Matches 298; Conservative 234; Mismatches 59; Indels 407; Gaps 63;
```

Qy	19	VKEISDKI-----SKEEMVRLKVVVTFTMDMQDSSEEEKLY-----	56
D b	1900	mkeidskhlqevqlmktkieacielekvlgelkkensdlsleklysfcdhgellqrsvts	1959
Qy	57	--LNLALHLASDFELHPGKDVRLLVACCLADIFRIVAPEAPYTPDVKLDKFITRQL	114
D b	1960	eglsndlemhad---ksredidgnva-----kvndswkerfidvenel	2000
Qy	115	KGLEDTSPQNRYFYLENIATWKSYNICFEEDSN--EIFTQYRTLFVSVINNGHNOK	172
D b	2001	srireskasiehealyleadlievqtcklekdnenkqvivcleeel-svtiserlql	2059
Qy	173	VHMHWVDLMSSICEGDPTVSQELLDTVLNVLVPAAHKNNKOAYDLAKALLKRTRQAATEPY	232
D b	2060	--rgeldtmskttaldqlslsekmlkek-tqeleshgscclhcigvaeaevekteqq--	2113
Qy	233	ITTFFNQVMILCKYSIDSLSHEVFDLI-----LELYNIDSHLLSLVLPOLPEFKLSNDNE	287
D b	2114	-----tlssdwseilkdktkhlqekglskdsqalsitkcelenlaqaqnke	2160
Qy	288	ERLOV--VKLLAKMFGAKDSLASONPLWCQYL--GRFNDIHPIRLECIVKFASHCLMN	343
D b	2161	kelllvkeseslqarlesdyeklrvskaealvekgef-----airfsscqeevhqlrr	2215
Qy	344	HPDIAKLDTLEYIKRVSDHPDEAIRHDVIYSTVTAACKDILLVNDDHLLNFVBERTLDKR--	401



obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures - Claim 13; SEQ ID 6892; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 92 AA;

Query Match 6.5%; Score 467; DB 21; Length 92;

Best Local Similarity 100.0%; Pred. No. 5.5e-24;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

839 MNHNSKSTSLRLTLTLHSDGDLTEGKISKPDMSRLRLAAGSAIVKLAQEPYHEI 898

|||||

1 mknhsksstslrltlhsgdglteggkiskpdmrlrlaagsaivklaqepcyhei 60

899 ITLEQYQALAINDECYQVROVFAQLKHK 929

|||||

61 itleqyqalaindecyqvrqvaqlkhg 91

# RESULT

AAW98879

IL AAW98879 standard; Protein; 1743 AA.

AAW98879;

31-MAR-1999 (first entry)

H. pylori GHPO 1755 protein.

GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis; peptic ulcer disease.

Helicobacter pylori.

WO9843478-A1.

08-OCT-1998.

01-APR-1998; 98WO-US06371.

29-JUL-1997; 97US-0902615.

01-APR-1997; 97US-0833457.

24-JUN-1997; 97US-0881227.

(HUMAN) HUMAN GENOME SCI INC.

(INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

WPI; 1998-542293/46.

N-PSDB; AAX14598.

New isolated Helicobacter polynucleotides - used to develop products for the diagnosis, prevention and treatment of Helicobacter infections and gastrointestinal diseases

Claim 8; Page 2000-2008; 2054pp; English.

XX

This sequence represents a Helicobacter pylori GHPO protein of the invention. The polypeptides can be used for preventing or treating Helicobacter infections and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

Sequence 1743 AA;

Query Match 3.1%; Score 225; DB 19; Length 1743;

Best Local Similarity 18.4%; Pred. No. 3.4e-06;

Matches 285; Conservative 217; Mismatches 523; Indels 520; Gaps 71;

QY 4 SKTRTN-DGKITYPGCVKEISDKISKEEMVRRLK-----MVVKT 41

Db 312 tetqtnfdgks-----eeltddndqelkgkkyiigivavlvilfsifby 366

QY 42 FMDMDQDS---EEKELYLNLAHLASDF--FLKHPGKDVRLLVACCLADIFRIYAPEAP 96

Db 367 fmpledksrfskdrnlvndeigirqeynrllkernekgnmid----- 410

QY 97 YTSPPKLDIFMFIYTRQLKGLDTSQPQPNRYFYLLNTAWKSYNICPELDSNEIFTQ 156

Db 411 -----knlff-----nddpnrtlyn1-nla-----eiedknpl----- 438

QY 157 LYRTLFSVINNGHOKVHMVMVMDLMSIIICEGDTVSQELLDTVLVNLPVPAHKNLKNQAYD 216

Db 439 ---rafyecisngngyeecikll-----kdkklqdg----- 466

QY 217 LAKALLKRTAQAIPIYITTFNQVLMLGKTSISLSEHFVDLILELYNIDSHLLSLVLPQ 276

Db 467 -----mkktlea-----yndcinknakt----- 483

QY 277 LEFKLSNDNEERLOVVKLLAKMFGAKDSELASQNKPLMQCYLGRFNDIHVPIRLECVKF 336

Db 484 -----eeerikcldli-----kdenl-----lnqgkvqvaldclkn 518

QY 337 A-----SHC---LMNHP-----DLAKDLTVLKVRSHPDEAIRHVDIVSTVTA 378

Db 519 aktdeerneckilndpeirekfrkelelqelqey-----kdciknaktea 565

QY 379 KKDILL--VNDHLLNFVRETLTLD---KRWVRKEAMMGLAQIYKKYALQ----- 422

Db 566 ekncklqiskeaierlkgaldclknaktdeernecklknipqdlqkelladmsvskaykd 625

QY 423 --SAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLIVERIFAQYMWVPHNLET--TERMKCLY 479

Db 626 cvskarnekeqec---eklltpearkklegqvl-----dciknaktddeerkkclkd 674

QY 480 LYATL--DLNAVKNALN-----EMWKNCONLL-----RHQVKDLIDL 512

Db 675 lpkdlqsdilakeslkaydcvskaqteaekkecelkltpeakilleeakesvaykdc 734

QY 513 IKPKRTDASVKAIFSKVMVITRNLPDQKADFMKFTQVLED-----DEKIRKQLEVL 566

Db 735 vsqakteaekkec-ekllt-----peakkkleeakskaylcvsvarnekekecelk 788

QY 567 VSPTCSCKQAQECVREITKLGONPKQPNPFLEMTKFLLERIAPVHIIDTESISALIKQVN 626

Db 789 ltpeak-killeqaldclknaktkerkclklpklqkv-----lakesvaylcvsv 843

QY 627 KSIDCTADDEGVPTDQAIRAGLELLKVLSTFHTPISFSAETTESLACLAMDDKVAE 686

Db 844 qakteaekkecelkltpea-rkllleak-----ksvkaxidc-----vsq 882

QY 687 AALQIFKWTGSKIEEDFPPIRSALLPVLHHSKKGPPRQAKYAI-----HCIHAIFSSKET 742

Db 883 akteaekkecelklt--pearkll-----eeakesvaykdcvskarneke 927

QY 743 QFAQIFELPHKSLDPSNLEHLITPLVTIGHTALLAPDQFAAPWKSVAFTFVKDMLMND 802















Region /note= "Hank's conserved region - beta strand 4"  
525..548  
/label= "Subdomain V"  
Region /note= "Hank's conserved region"  
525..532  
/label= "Beta strand 5"  
Region /note= "Al[ha helix D"  
540..548  
/label= "Al[ha helix D"  
Region /note= "Subdomain Via"  
554..567  
/label= "Hank's conserved region - alpha helix E"

WO200050454-A1.

'31-AUG-2000.

24-FEB-2000; 2000WO-US04732.

24-FEB-1999; 99US-0121461.

(TUFT ) TUFTS COLLEGE.

Soto AM, Sonnenschein C, Geck P, Szelei J;

WPI; 2000-565451/52.

N-PSDB; AAA28051, AAA28052.

New human androgen-induced tumor suppressor cDNA sequence termed

'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient  
diagnosis and treatment of prostate cancer -

Claim 2; Fig 1; 152pp; English.

This invention relates to a human androgen-induced tumour suppressor cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell proliferation and use as a marker for the efficient diagnosis and treatment of prostate cancer. The invention includes AS3 cDNA and protein sequences, a vector comprising the cDNA sequence, a host cell transfected with the expression vector, and a method for producing an AS3 polypeptide comprising culturing the transfected cells. AS3 has cytostatic activity, and acts to suppress cell proliferation. The AS3 gene is useful as a marker for the efficient diagnosis and treatment of prostate cancer. The AS3 nucleic acid molecule can be used as a source of antisense agents for sequence specific modulation of gene expression. The AS3 protein may be used in the treatment of disorders caused by aberrant modification or mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene or aberrant post-translational modification of the AS3 protein. This sequence represents the human AS3 protein sequence.

Sequence 1391 AA;

Query Match 100.0%; Score 7193; DB 21; Length 1391;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MAHSKTRTRNDGRTYPPGKESIDSKSEEMVRRLLKVVYKTFMDMDQDSEEEKELYNLIA 60  
1 mahsktrtrndgkityppgkveisdskiseemvrrlkmvvtftmdmdqdsseeekelynlia 60  
61 LHLASDFFLKHGKQVRLVACCLADIFRIYAPEAPYTPSPDKLKDIFMPTITRLQKLEDT 120  
61 lhlasdfflkhpqkdvrlvavaccladifriyapeapypstpsdklkdifmptitrlqlgledt 120  
121 KSPQNRXYLLNTAWKSYNICFELEDSNEIFTOLYRTFLSYVNNHNOKVHMHVVDL 180  
121 kspqnrxyllntawksynicfeledsneiftolyrtflsyvnnhnokvhmvvd 180  
181 MSIIICEGTVSQELDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQALIEPYITTFNFQV 240  
181 msiiicegdtvsqelldtvlnlvpahknlnkqaydlakallkrtaqaaliepyitffnqv 240

QY 241 LMLGTSISDLSSEHVFDLILELYNIDSHLLSVLPQLEFKLASNDNEERLQVVKLAKNF 300  
DB 241 lmlgtsisdsissehvfdlilelynidshllsvlpqlefkllksndneerlqvkvllakmf 300  
QY 301 GAKDSELASQNKPLWQCYLGRFNTHVIRLCEVAFASHCLMNHDPDLAKDLEYLKVRSRSH 360  
DB 301 gakselasqnxplwqcyigrfnthvirlecvaafashclmmhdpdlakdleylkvrsrh 360  
QY 361 DPEEAIHRHDVIVSIVTAARKOILLVNDHLLNFVRETRTLDKRWVRKEAMMGLAQYKYKA 420  
DB 361 dpeearihdvivsiytaarkoillvndhllnfvertldkrwvrkeammglqaykyka 420  
QY 421 LOSAAGKDAAKOIAWIKDKLLHIYYQNSIDDRLLYVERIFAQYVWPHNLETTTERMKCLLYL 480  
DB 421 lqsaagkdaakqlawikdkllhiyyqnsidrrlllyverifaqymvphnlettermcklyyl 480  
QY 481 YATLDNAVKALENMVKONLRLHOVKDLIDLIKOPKPTDASVKAIFSKVMVITRNLDPDG 540  
DB 481 yatldnavkalenmwkcqnlrlrhvqkdldlikpkrtdasvkaifskvmvitrnlpdpg 540  
QY 541 KAQDFMKKFTQVLEDEKIRKOLEVLVSPSCCKQAEGCVREITKKLGNPKQPTNPFLM 600  
DB 541 kaqdfmkkftqvledekirkolevlvspscckqaegcvreitkklgnpkqptnpflm 600  
QY 601 IKFLLERTAPVHIDTESISALIKQVNSIDGTADDEDEGVPTDQAIRAGLELLKVLFSFTH 660  
DB 601 ikfllertapvhidtesisalikqvnksldgtaddedegvptdqairaglellkvlfsfth 660  
QY 661 PISFHSATFESLLACLKMDDEKVAEALQIFKNTGSKTEEDFPHRSALLPVLHHKSKK 720  
DB 661 pisfhsatfesllaclkmddekvaeealqifkntgskteedfphrsallpvlhhsksk 720  
QY 721 GPPROAKYAIHCHIAIFSSKETQFAQIFELPKSLDPSNLEHLITPLVTIGHIALAPQ 780  
DB 721 gpprqakyaiahchiaifssketqfaqifeplksldpsnlehlitplvtighiallapq 780  
QY 781 FAAPKWSVATFIVKDLLMNDRLPKKTKLWVPDEEVSFETWIKVQAIKMMVRLGLMK 840  
DB 781 faapkwsvatfivkdlmndrlpgkttklwvpdeevsfetwvkiqaimmvrvlglmk 840  
QY 841 NNHSGSTSTRLTLTILHSDGDLTEQGRISKPDMSRLAAGSAIVKLAQBPCEHEIT 900  
DB 841 nnhsgststlrlttilhsdgdltteqgiskpdmsrlaagsaivkllaqbpceyheit 900  
QY 901 LEQYQCALAINDECYQVRQVFAQLHGLSLRLPLLEYMAICACAKDPVERRAHARQ 960  
DB 901 leqyqlcalaindecyqvrqvfagqlhglslrlpleymaicalcakoverraharq 960  
QY 961 CLVKNINVRREYLKQAAVSEKLLSLPEYVVPYTHLLAHPDQVYKQDIEQLKDVKEC 1020  
DB 961 clvkninvrreylkqaaavsekllslpeyvvpvpythllahdpdvvykqdieqlkdvekc 1020  
QY 1021 LWFVLEILMAKNENNSHAFIRKVENIKQTKDAQDPDDAKMNEKLYTVCDAVNNIIMSKS 1080  
DB 1021 lwfvleilmaknennshafirkmvenikqtkdaqdpddakmnekytvcdavnniimsk 1080  
QY 1081 TTYSLKSPKDPVLPARFFTPQDKNFNTKYNLPPENKSEFTPGKPKTKTNVLGAVNKLPS 1140  
DB 1081 ttylskspkdpvlparrfftpqdknfntkynlppemksftpgpkpkktnvlgavnkplss 1140  
QY 1141 AGKQSQTSRRMETVSNASSNSPSPGRIKGLDSEMDHSENEEDYTMSSPLPGKSKDK 1200  
DB 1141 agkqsqtssrrmetvsnassnspspgrikgrldssemdhsenedytmssplpgkksdk 1200  
QY 1201 RDSDLVRSLEKPRGRKKTPVTEQEKLGMDLTKLVQEQPKGSQRSRKRGHTASED 1260  
DB 1201 rdsdlvrslekprgrkktpvteqeklgmddtklvqeqpkgsqrskrghtasesd 1260  
QY 1261 EQQWPEEKRLKEDILENEDEQNSPKKGRGRPPKPLGGGTGPKKEPTMTKSGSKSKKSG 1320  
DB 1261 eqqwpEEKrlkediLENEDEQNSPKKGRGRPPKPLGGGTGPKKEPTMTKSGSKSKSG 1320  
QY 1321 PPAPPEEEEEERQSGNTEQKSKSKOHRVSRRAQORAESPESSAIESSTQSTPQGRGRPSK 1380











3313 HGI-----DAMLLNIYHYFEQFLSKVSEAIHQKSLNSIKERILLMSWKDT-- 3360  
408 AMGLAQIYKVKALQSAQDAKAQIAWKDKLHLHYIYQNSIDRLVERIFAQYVPHN 467  
3361 -----NVALKESAKSHALF-----KVLHYRE-----VLQPVSSYLSQKHD 3400  
468 LETTERMKCLYLYATLDLNAVKALEMW--KQCNLLRHQV-----DLDLLIKQPKTDAS 521  
3401 WD-----SLDITENNSA---MVAKKVNLSPSYIEKMDTEIMKLV--PVRFEN 3443  
522 VKAIFSKVMVITRNLPDPK-----AODPMKFTQVLEDE-----KIRKO 562  
3444 TPTTLRLMTLFAVKEPGSTETNMVSNLITDARELMKLTPTETINDNLSEIKHLSRKH 3503  
563 LEVLVSPTCSQKQAEQCVREITKLGKNGKOPTNPFLEMKIFLLERAPVHIDTESALI 622  
3504 L--LLTETFTKLFAGLQVRV--KAGIEENLSN-----LRLNLLAVIPTPVTSLSIEKVD 3554  
623 KOVKNKSID-----GTADDEDEGYPTDQAIKAGLELLKVLSTFTHTPISFHSAE-----T 669  
3555 RSLMKSLDFPKFOTLAGHQHNDLSP--EVOKGVGL-----FNSMLSIQLGERAQLVE 3606  
670 FSLSLACLK--WDDEKVAEAAQIFKNTGSKTEEDPPHRSALLPVHLHKKSGPPRQAK 727  
3607 FTNELALKNVYSEVGVNGSPLESEFNNS-----SPNEVSS--LGYDHDHFENRA--QAV 3655  
728 YAIHCIAHAFSSKETQFAIFEPLHKLSDPSNLEHLITPLVTIGHIALLAPQFAAPWKS 787  
3656 SMCQIYIAVIOK-----HSSISPT-----ASFQSIGH-----ELSRFA----- 3689  
788 WVATFIVKDLMLNDRLPG-----KTTKLWVPDEEV 818  
3690 -----DLSNKLFPSSIPYASADKYSSIRDOQKINGINDLIEYCRKRTTEL--PELSY 3739  
819 SPETWYKIOAKMMWRWLGKNNHKSSTSTLRLTTLHSDGDLTEOGKISKPDMSRL 878  
3740 CFKHLVSLQSLKSISRTQDLTND-----EFLNLMFNVLNLFDSLSSTETATKNMRTF 3793  
879 R-LAAGSALVKLAQEPYHEITLEQYQCALAINDECYQVQVFAQKHLKGLSLRLP- 936  
3794 KELAETSSTFIESS--CFSKVLRAFNLKQSMKLSLSEKLSKSSVDKMSQCLLMFLFVP 3851  
937 -----LEY-----MAICALCAKDPVKERRAHAROCLVK 964  
3852 CEQFINLAESVLDYFINVHNSNLSLSKISTLFFWMYANNGFCSPDLPOEGSKNSGE--- 3907  
965 NINVRREYLKQHAUV-----SEKLSLLPEYVYPTIHLHADDPVKVQDIE--OLKDV 1017  
3908 -----LESCTGLSGVGVAEDITWLA--DDDLEELANEEDTANGSDLDESEAREL 3956  
1018 KECLWFVL-EILMAKNENSHAFIRKVENIKOTKDAQG---PDDA--KMKELYTVCDV 1071  
3957 ESDMGVTKDSVVSENESSDS-----BEENODLDEEVNDIPEDLSNLEKLV----- 4004  
1072 AMNIIMSKSTYTSLESKDPVLPARFPTQPKNFSTKNYLPPEKMSFETPGKPTTNVL 1131  
4005 -----DEP-----NEEDLLETEQS-----NEQ 4022  
1132 GAVNKPLSAGKQSQTKS-----SRMTVNASSSSPSPGRIKGL--DSSEMDHSHENED 1186  
4023 SAANNESDLVSREDDNKALDEKDRQEKDEEEMSDVGIDDEIQDQENNSOPPPENED 4082  
1187 YTMSSP---LPGKSKDRDDSLVSELEKPRGRKKTPTVEEKLGMDDLTKLVQEQK 1242  
4083 H-LDLPEDLKLDBKGDVSKDSLEDMDMEADENKEEADAEDKEDP--MODFEDPLEENN 4139  
1243 PKGSQRSRKRGHTASDESQOMPEEKRLKEDILENEDEQNSPPKKGKRGPRPKPLGGGTP 1302  
4140 TLDEDIQDDFSLADEDEKM--NEDGFENYQENESTEDGVKSDDELEQ-----GEVP 4192  
1303 KEPTWTKSKKSGKSGPPAPEEEEEERQSGNTQKSKSKOHRVSRRAQRAESPSS 1362  
4193 EQDAIDNHPKMAKSTFASAEADEENTDKGIVGENEELGEDCAESGVGRNGTADGEFS 4252

QY 1363 AIESTQSTPQKGRGRPSKTP 1382  
DB 4253 SAEQVQ-----KGEDTSTP 4266

RESULT 15  
T18296  
myosin heavy chain - Entamoeba histolytica  
C:Species: Entamoeba histolytica  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000  
C:Accession: T18296  
R:Guillen, N.  
submitted to the EMBL Data Library, February 1997  
A:Reference number: Z18865  
A:Accession: T18296  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2139 <GUI>  
A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1  
C:Genetics:  
A:Gene: mhca  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.1%; Score 223; DB 2; Length 2139;  
Best Local Similarity 19.6%; Pred. No. 0.0086;  
Matches 292; Conservative 233; Mismatches 570; Indels 396; Gaps 62;  
QY 2 AHSKRTNDGKI-TYPPGVKEISDKISKSEEMVRR-LKVVVKTFTMDQDSEEEKELYNL 59  
DB 894 ATAKTGELEAKTDLEDKISELSAAELDKQELNLAIE---NLEEDKEELKETIDNL 950  
QY 60 ALHASDFFLKHGPKDVRLLVACCLADIPRIYAP-----EAPYTSPPKLDIPMFIT 111  
DB 951 KGLDKDS---KLKGEDLEVEITELNSQINTLNATVBDKDKTAAEMQESIDEKEDETKLK 1007  
QY 112 ROLKGLGLEDTKSPQFNRYFYLEENIAWV-----KSNICEFEEDSNEIFTQLYRTLF 162  
DB 1008 GDIKLLEEKEDD-----LEQDRADVSAKDDIAKLNKITECEDAKDEAKLEOLE 1060  
QY 163 SVINNGHN-----QKVHMHVMDLMSIIC-----EGDTVSQELDITVLVNLVPAH 207  
DB 1061 DEENKKNKDLTNELOQTLKLGTEKSLAAQVAATKKAASDERDTLSQNLNLEKLT----- 1115  
QY 208 KNLNKOAYDLAKALLARTQAITEPYITTFNQVLMGLK7SISDLSEHVFDLILELYNIDS 267  
DB 1116 KNLTKTKADLEK-----KISGLKQDYEDL-----EDDK 1143  
QY 268 HLLLSVLPQLPEFKLSKNDNEERLOVVKLLAKMPGAKDSELSQNKPLWQCYLGRFNDIHV 327  
DB 1144 NKIEGDLRNAQRKIKELDDE---ITK-----GADVSYLOKQKEEYESQIAKM----- 1188  
QY 328 PIRLECVKFAHCLMHNHPDLAKDLTEYLVKVRSHDPEAIRHVDIVSIVTAAKKDILLVND 387  
DB 1189 -----QEEKEATGNVKNKEKTIKEKELEIQS- 1215  
QY 388 HLLNFVRETLTKRWVRKRAMGLAQIYKKY--ALQSAAGKDAKQIAWIKKLLHIYYQ 446  
DB 1216 -----LQEKLDTEVEKEDEAKKKKEIEKEMKALQEEKNVENESSKSTKDK----- 1262  
QY 447 NSIDDRLLVERIFAQYVPHNLETTERMKCLYLYATLDLNAVKALEMWKNLRLHGV 506  
DB 1263 KKLED-----NLKDTQ-----KKLDDMTADNEKLKAKA 1290  
QY 507 KOLLDLIKOPKTDASVKAIFSKVMVITRNLPDPGKADPFMKFTQVLEDEKIRKOLEVL 566  
DB 1291 KDLAQLNEVO--DNHEKAV-----ADALELNKKAQSKDELNSLKAELEAL 1335  
QY 567 VSPTCCKQAQECVREITKLGKNGKOPTNPFLEMKIFLLERAPVHIDTESIALIKQVN 626  
DB 1336 T-----KAKSVESKNDSENEKAALS---BEIDQANEKLNKIQADLRKATADLQEAN 1385



623 LGVVTAPTSTNKNLSTDMFELFWHLITPIYEDVYFKYDSIHLQNVLTFTVLHLLIGKNA 682  
QY 409 MGLAQIYKKY-----ALQSAAGKDAK---QIAWIK-----DKLLHIYQNSIDRLLV 455  
DE 683 DVALERKYYKHHMPSVITASBGVLKDISSLPPOI--IKREYDKIMKVFEQT-----V 733  
QY 456 ERIFAQYVMPHNETTERMKCLYLYATLDLNAVAKALNEMKQCNLLRHQVKKDLDLLIKO 515  
DE 734 EVAISNVNLAHDLITS-----LKHLPED----- 757  
QY 516 PKTDASVKAIFSKVMVITRNLPDPCAKQDFMKKFTOVL-----EDEKIRKOLE 564  
DE 758 RKDQTHLESFSLILKVTQNNKDPIDPFCAGVTSFVYTFDLFLRKNDSLVNFIQ 817  
QY 565 VLVGPTSCCKAEGCVREITKGLNKPQTPNPFLEMIKFL-----ERIA 609  
DE 818 ISKVGISQGNMTDLLKDVIRKARN--ETSEFLIEKFELEDDDKTEVYAQNVMVGSTLL 874  
QY 610 PVHIDTESISALIKOVKNSIDGTADDEGYPTQAIIRAGLELLKVLSTFTPIFHSNET 669  
DE 875 PNISFREFOQLANIVNR-----VPNENSIENFLDLCLKLSF--PVNL----- 915  
QY 670 FESLIACLKMDKEVAEALQIFKNTGSKIEEDF-----PHIRSALLPVLHHS 718  
DE 916 FTLLHVSWMNNFIYFQSYVSKNE-NKLAVDLITLLKSLPGNPELFSGLLPFL----- 970  
QY 719 KKGPPROAKY---AIHCIIHA-----IFSKKE- 741  
DE 971 -----RRNKFMIDILEYCITHSNPNLLNSIPDLNSDLLKLLPRSRASYFAAKIKLFCSEQ 1025  
QY 742 -----TOFAQIPEPLKSLDPSNLE--HLITPLVTGHTALLAPODFA 782  
DE 1026 LTLVRWLKGOOLEQNLQNFSEIENVLQNASD-SELEKSEIIRELL---HLAMANPIE-- 1079  
QY 783 APKWSVATFTVKDOLLN--DRLPCKTKTKLWVPDEEVSPETMVKIOAKMVMRWLLGMK 840  
DE 1080 -PLFSGLLNFCIKNMADHLDPCGNTSEVLF---KISPELLIKLLTYK----- 1125  
QY 841 NNHSGKSTLRLTLTLHSDG-----LTEQGIKSKPDMRSLRLAAGSAIVKLAQEPCH 896  
DE 1126 --EKPNKLLAAVIEKIEGNGDDVILELEKIIIOKE-----IOILEKLKEP--- 1170  
QY 897 EIITLEOVOLCALAINDECYO-----VROVFAQKLHGLSLRPLPLEYMAICALCAKDP- 950  
DE 1171 ----LLVFFLNPVSNMOKHKSNMURELVLLYTLPLSRSAKAKFFSMILISILPPNP 1226  
QY 951 -----VKERRAHARQCLVKNINVRREYLKQ-----HAAVSEKLLSL 986  
DE 1227 YQIDMVNLLIDLKSHNRKFKRXY--NATLKTIG--KWIQESGVVHOGDSKEIEA 1281  
QY 987 LPE-----YV-----VPYTIHLAHDPPYKVQDIEQLKDVKEC 1020  
DE 1282 IPDTKSMYIPCEGSENKLSNLQKVDSODIQVPAT-QGMKEPPSSQISSQISAKD-SDS 1339  
QY 1021 LWFVLEILMAKNENNSHA-----FIRKVMENIKQTK-----D 1052  
DE 1340 ISLANTAIMNSQOESHANRSRIDETPLEVDNESIREIDQOMKSTOLDKRVANHSNIC 1399  
QY 1053 AQGPDPAKMKNEKLTVCVAMNIMSKSTTYSLESPPKDPVLPARFFTPDKNFNNTKNYL 1112  
DE 1400 STKSDEVDVTE-LHESIDTQSSEVNAVQIEVLITSELKAVTNRSIKTNPDHNVVNSDNL 1458  
QY 1113 PPEKSFPTPKPKPT-----TNVLGAVNKPLSSA----- 1141  
DE 1459 KRPSKETPTSENKRSGHETMVDVLVSEQAVSPSSDVICTNIKRSIANEESLALRSIK 1518  
QY 1142 -----GKQSTKSSRMETVSNASSSSNPSSPGRIKGLDSS----- 1178  
DE 1519 VETNCNENSLNVTLDLQDTITKEDGKQVHEVORQENQESMNNKINSKSFQODNIAQVKS 1578  
QY 1179 -----MDHSENEVDYMSPLPGKSKDRDSDLVRS-----LEKPRGRKKTVPVTEOE 1227  
DE 1579 VKARPNNEGENDYACV-----EQASPVREVPDGIQIPSGTILLSSKQTE 1628

QY 1228 KLGMDDLTK-----LVQEOKPKGSQRKRGRHTASEDDQMPPEKRLKEDILEN--- 1277  
DE 1629 KSKYDDLRSDEDEHGTVAQKHQVGAINSRNKNDRMDSTPIQGTTEESREVVMTEGIN 1688  
QY 1278 --EDEQNSPPKKGKRGPRPKPLGG-----TPKEPTMTKTKSKGSKKSG 1320  
DE 1689 VRLDESGTCELNKLKG-----PLAGDKDANINDDFVPVEENVROEGFLKSEHAVSKETG 1744  
QY 1321 PPAPEEEEEEE-----ROSGTEQKSKQHRVSRRAQOARESPSS 1362  
DE 1745 -----LEEQPEVADISVLPEIRIPFNLSLKMKGSKQIKELKRLQRLNELMPDPPRM 1799  
QY 1363 AISTQSTPQKGRGRPSKT 1381  
DE 1800 T-ENTNINAQNGLDTPKPT 1817  
RESULT 14  
T41581  
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: T41581  
R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z22002  
A:Accession: T41581  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4717 <MUR>  
A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08  
A:Experimental source: strain 972h(-)  
C:Genetics:  
A:Gene: SPDB:SPCC737.08  
A:Map position: 3  
C:Superfamily: Saccharomyces probable membrane protein YLR106C

Query Match 3.1%; Score 223.5; DB 2; Length 4717;  
Best Local Similarity 18.4%; Pred. No. 0.023;  
Matches 290; Conservative 249; Mismatches 572; Indels 469; Gaps 71;

QY 24 DKISKEEMVRLLKVVTFMDMDQDSEEEKELYLNALHLASDFLKHGPKGKDVRLVACC 83  
DE 2935 DVDS--NLIEQIKV-----LDLNDISIRFEIYFLASRLCSE---KQHSDDTH---S 2977  
QY 84 LADIFRIYAPEAPYTPDKLKDIFMTITROLKGLDTSKQFNRYFYLLLENIAWVK---S 140  
DE 2978 LANSFVLNANEF-YIHNAKIKQ-----KELEETEKNRLYRQREFNFDKN-DYLKVFIN 3029  
QY 141 YNICFELEDSNEI-----FTQLYRTLFVSINNGHNQKVVHMHVMDLMSI-----I 185  
DE 3030 YDDEVEPEVEPEVIERKRFLOQAFWLSLYNEIYSEKMNVIPLQMLNTGSLAKKIKV 3089  
QY 186 CEGDTVSOELLDTVLVNLVPAHKNLKOAY-----DLAKALLR 224  
DE 3090 KNPDMIASSGPDIVSVVLMGMVKSTNERQYWTTPVYFNYPNPSKATEVRDLKIVESR 3149  
QY 225 TQAIEPIVITTFNQVLMGLKTSISDLS-----EHVFDLILEL-----YNI 265  
DE 3150 AISLIKWPENFVIRGLKDAIDAILNSPFSPIAEYLSKLERVHLLSEWEKLSRESYSL 3209  
QY 266 DSHLL---SVLPOLERKLSNDNEERLQVVKLLAKMFGA-----KDSLEA 308  
DE 3210 ANEMDLIKKIIDMRKPELSNMNLLKLEEVKLSERVYPRLYSILOFIIKPFPPENSFT 3269  
QY 309 SQNPLWQCYLGRF-NDIHVPRLCYKVFASHCLMNHDPDLAKDLTEYLKVRSHDPEAIR 367  
DE 3270 KQNLCEASIIQVITDITLVGEFQCLK-----CLLSFSQHAASL-----RIC 3312  
QY 368 HDVIVSVITAAKKDILLVN-----DHLLNFVRERTLDKRWVRKE 407



```
QY 720 KGPPOAKYAIHIAIFSKSTQFAQIFELHKSLOPSN-----LEHLITPLVTIGHIA 774
DB 1011 KVVGDIDKODLEAKNT-QTLTADVDKNEIEBOQRKIFSLIQEKNELOQMLESVIAKEQL 1069
QY 931 KDDLKQLESQI-----ERDLKSDIHDVTVMNIDTQELRNALLESKQHOETINTL- 983
QY 775 LIAPDOFAAPKSWAFVIFVOLLMDRLPGKTKLWVPDEVSPTVVKVQIAKMMVYR 834
DB 984 -----KSKISEEYSRLHM-----BENTGETKDEFQ-----Q 1010
QY 835 WLLGMKNHKSSTSLRLTLILHSDGLTBOGKISK--PDMSLRLAAGSAIVKLAQ- 891
DB 1011 KVVGDIDKODLEAKNT-QTLTADVDKNEIEBOQRKIFSLIQEKNELOQMLESVIAKEQL 1069
QY 892 -----EPCYHEIITLBOYLQALAINDECYQVQVFAQKLHKLGLSLRLPLEYMAICALC 946
DB 1070 KTDLKENIEMTIENQELRL-----LQDELKQOEIVAQEKNAHKK-----EGELSRFCDR 1122
QY 947 A--KDPVKERRAHARCLVKNINVRREYLKQHAAYSE-----KLLSLLEPYVVPYIT 996
DB 1123 AEVEEKLKESQOLQEKQOOLLNVQEMSEMOKKINEIENLKNELKNKELTLEHMETERL 1182
QY 997 HLLAH--DPDYVVKVODI--EOLKDVKECLWFVLEILMAKNENNSHAFTRKMVENIKOTKDAQ 1054
DB 1183 ELAQLKNEYEVKSTKERRKVLKE-----LQSPETERDHLRGYIREATGLQTKREL 1237
QY 1055 GPDDAKMNEKLYTVCDAVMIIMSKSTTYSLESPKDPVLPARFFTPQDKNFSNTKNYLP 1114
DB 1238 KIAHILKHEQETIDELRSV--SEKT-----AQIINTQDLEKSHTK--LQE 1280
QY 1115 EMKSFTPGKTKTNVGLAVNPLSAGKQSTKS--SRMETVSNASSNSPSSPGRIKG 1172
DB 1281 ETP-----VHHEEQELLNVKVSQETQTMNELELLITEOSTTKDSTTLARI-- 1326
QY 1173 RLDSSMDHSE--NEDYTMSSPLPGKSKDRDSDLVRSLEKPRGKKTPTVE----- 1224
DB 1327 -----EMERLNRKEFQESQEKSLTKERDNLKTIKEALEVKHDLKHEHIREITAKIOE 1381
QY 1225 -----QBEKLGK-----DDLTKLVQEQ---RKP-----GSRQSRKRGH-----TA 1256
DB 1382 SOSKQBSLNMKEKQNETTKIVSEMPQFPKDSALLRIEIMGLSKRLQESHDEMKSYA 1441
QY 1257 SPSDEQ-----WPEKRLKEDI-----LENEDEQNSPPKKGKRRGPP--KPLGGGTP 1302
DB 1442 KEKDDLQRLQEVQESDQLENKIBVAKHLETEBELVAHCCLEKEQETINELRVNLS 1501
QY 1303 KEETPMKTSKK-----GSKKSGPPAPEEEEE---EROSGTEQKSQKOHVSPRAQ 1354
DB 1502 EKETEISTQKOLEALNDKQNKIQUIYEBEQOLNIKQISEVOENVNLEKQKERRKAKD 1561
QY 1355 RA-ESPSSAIEST 1367
DB 1562 SALQSTESKMLEIT 1575
RESULT 13
S46157
R1P1 protein - yeast (Saccharomyces cerevisiae)
N1Alternate names: protein YBR1743; protein YBR275c; RAP1-interacting factor
C1Species: Saccharomyces cerevisiae
C2Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
C3Accession: S46157; S46156; S44537; S25704; S39131
C4Brandt, T.; Christiansen, C.; Holmstrom, K.; Kalliesoe, T.
submitted to the Protein Sequence Database, August 1994
A1Reference number: S46157
A2Accession: S46157
A3Molecule type: DNA
A4Residues: 1-1096 <BRA>
A5Cross-references: EMBL:Z36144; MIPS:YBR275c
A6Experimental source: strain S288C
R1Aigle, M.; Baclet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to the Protein Sequence Database, August 1994
A1Reference number: S45940
A2Accession: S46156
A3Molecule type: DNA
```

```
A:Residues: 1005-1916 <AIG>
A:Cross-references: EMBL:Z36144; MIPS:YBR275c
A:Experimental source: strain S288C
R:Holmstrom, K.; Brandt, T.; Kalliesoe, T.
yeast 10(Suppl.A), S47-S62, 1994
A:Title: The sequence of a 32420 bp segment located on the right arm of chromosome II
A:Reference number: S44537; MUID:94378722
A:Accession: S44537
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1096 <HOL>
A:Cross-references: EMBL:X76053; NID:g600025; PID:g429120
A:Experimental source: strain S288C
R:Hardy, C.F.J.; Sussel, L.; Shore, D.
Genes Dev. 6, 801-814, 1992
A:Title: A RAP1-interacting protein involved in transcriptional silencing and telomer
A:Reference number: S25704; MUID:92249772
A:Accession: S25704
A:Molecule type: DNA
A:Residues: 1-579,'S',581-731,'A',733-1916 <HAR>
A:Cross-references: EMBL:X66501; NID:g4327; PIDN:CAA47121.1; PID:g4328
C:Genetics:
A:Gene: SGD:RIF1
A:Cross-references: SGD:S0000479; MIPS:YBR275c
A:Map position: 2R
C:Superfamily: Saccharomyces cerevisiae RIF1 protein
C:Keywords: transmembrane protein
F:263-279/Domain: transmembrane #status predicted <TM1>
F:296-312/Domain: transmembrane #status predicted <TM2>
F:438-454/Domain: transmembrane #status predicted <TM3>
F:663-679/Domain: transmembrane #status predicted <TM4>
F:788-804/Domain: transmembrane #status predicted <TM5>
F:905-921/Domain: transmembrane #status predicted <TM6>
Query Match 3.2%; Score 227; DB 2; Length 1916;
Best Local Similarity 17.4%; Pred. No. 0.0048;
Matches 316; Conservative 267; Mismatches 612; Indels 624; Gaps 80;
QY 3 HSKTRTNDGKIY-----PPGVKEISDKI-----SKEEMVRRLKMW--- 38
DB 183 YKTRVNPRALEYVSGETHGLVDNESVEFKEIEGGLILRQESDYVARRFEVYATF 242
QY 39 -----VKTFMDQDQSEEEKELYLNLA-----LHLASDFFLKHGPKDVRLLVACCL 84
DB 243 NLIIPILTTKKNVNDQKFNILIVNIESIEICIPHQLAQDTLLSSSEKKNPFVI--- 298
QY 85 ADIFRIYAPEAPYTS-----PD---KLKDF-----MFTR 112
DB 299 -----RLYVQIVRFFSAIMSNEFKIVKWLTKRPDLVNLKVIYRWTTGALRNSNKIIITA 354
QY 113 QLKGLDTSKPOFNRYFYLLNI--AMVKSNTCFELEDSENEIFTQLYRTLFSVINGNHQ 171
DB 355 QVSFLRDEK---FGTFPLSNEEIKPIITFTTEINSHNLIYEKL-----LLIRGFLS 405
QY 172 KVHMHMVDLMSS-----IICEGDTYSQELDLVNLVPAHKNLNKQAYDLAKAL 222
DB 406 KYPKLMIEVTISWLPGEVLPRIIGDEIYSMKILITSIVLL-----ELL 450
QY 223 KRTAQAEPIYITTFNQVLMGLKTSISDLSERFVDLILELY---NID-----SHLLSVLP 275
DB 451 KKCLDFVDEHERIY--QCIMLSPVCETIPEKFLSKPLANSYDSANLKDVKYTIHLLTQQIK 508
QY 276 QLEFLKLSNDNEERLQV--VKLLAKMF--GAKDSELASQNKPLW-----QCYLGRFRNDIIV 327
DB 509 --NYIVKNDKNIAMDLWSMTGLLYDSGRVYDLTSESKNVWFDLNLCLFINN-----HP 562
QY 328 PIRLECVK---FASCL-----MNHPLDAKLDTXYLK-----VRSHDPEAIRHDV 370
DB 563 KYRLMSIKVWRITTYICTKISQKQOEGNKSLSLLRTFQMTLPVYNDFPSAREGIIYHL 622
QY 371 IVSIVTA--AKKDIL-----LVNDHLLNFRERTLDKRWVR---KEA 408
A:Molecule type: DNA
```



943 LSDSKYKAEFAMVKVLKEONSERIASLESELNSLAVKERCESKLSGEVEHLKRO 1002  
QY 629 IDGTADDEGV-----PTDQAIRA--GLELLKVLSTFHPISFSAET 669  
DB 1003 LDDSSQKHKEALAKNIEIKOLINAKKATSDLAIKSEMGAOLQKAVD-THKSEFSAON 1061  
QY 670 PES-----LLA-----CLKMDKVAEAL----- 689  
DB 1062 ELSRSLDLALKEGEVERLNKEAALROBEIOOOOQTITKLEETEALAAALDKVALOEKE 1121  
QY 690 ---OIFKNTGSKIEDPHIRHALLPVLLHHSKK-----GPPRAKYAIHCHIAIFSSK- 740  
DB 1122 IKQVQATKGA--EKEMAKLAS--VISEKSKRIECLDQIQNRDLSICQIQHOSKL 1175  
QY 741 -ETQFAQ-----IFEPLHK-----SLDPSNLEHLITPLVTIG 771  
DB 1176 GESQGLQALJADLEKKKEOKELICEAQNAKAAEAKTLASEKASVSEKQLEGIOALEIEIG 1235  
QY 772 HIALAPD---QFAAPW-----KSWATFIVKDLIMN-DRLPQKKT-----KLWVP 814  
DB 1236 KERQACDQLOKLESLMAVOBEKETELQALKKELFKVQOELEQSQTSFTDSSGEALLYLS 1295  
QY 815 DEEVSPETMVKIOAKMVMRWLLGKMKNNHSGSTSTRLTLTILHSDG-----LT 865  
DB 1296 EAQERQALFAKEQAEQYKEIEHMKNEVNSLQAEIKILSSKVTTNNEVSVDFEQRLLK 1355  
QY 866 EQGKISPD---MSRLRLAAGSAIVKLAQPCYHEIITLEQYOLCALAINDECYQVQVFA 923  
DB 1356 ETSKSALEKMKQHLMELEASFRELLEKNCALDCLTEAQLAGEA--DOORMAVDSLQ 1413  
QY 924 QKL-HKGLSRRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAASEK 982  
DB 1414 QKLSKAEHTNHTLOQEIQAOMKNAE---KEQQICSLQ---QNLKSNQSLLEEFASLKH 1467  
QY 983 LLSLPEVVPYTH---LLAHPDYVKVQ--DIEOLK-DVKECLWFVLEILMAKNENNSH 1037  
DB 1468 YOEIARLDQKQKQBELSHKLLTERFQAEKAKEDMTEIV--LUKEKLHNOELQLH 1525  
QY 1038 AFIRKVENIKQTKDAQPDQAKM--NEKLYTQCDVAMNIIMSTYTSLESKPDVPLA 1095  
DB 1526 KFOSENSVSLTOISHLQOVNSOLLGANGOSLSOISDOGAKKLESEMSTLK-BOHKE----- 1579  
QY 1096 RFTQPKNFENTNLYLPPEMKSFETGCKPKTNTVGLGAVNKPLSAGKQSKTSRSMETV 1155  
DB 1580 -----EMKTLRL-----QYEKTLREGNKQVQETSLOLETV 1609  
QY 1156 SNASSSSNPPSGRIKRLDSDSEMDHSENEYDTWSSPLPGKKDKRDDSDLVRSLEKPR 1215  
DB 1610 -----TSKYDH-----VSKSVLKDQ 1624  
QY 1216 GRKKTPTVQEEK-----LGMDDLTK-LVQEQPKGQSRQ---KRGTASESDQOQWPEE 1267  
DB 1625 -----KTFQEKQRLLLQVQELNKLQSQEKTIRSQOQKLQREGTHEEADK-----SH 1674  
QY 1268 KRLKEDILEDEQNSPPKGRPRPKPLGGTTPKEPTHTKSKSKSKSGPPAPEEE 1327  
DB 1675 KRVLK--LESQLEQOOTQAVEHYKAQMEK-----AKVHYDAKKQON-----QE 1714  
QY 1328 EEEERQSGNTEOKSKSOHRVSRRAOQRAESPE 1360  
DB 1715 LSELQSHIKQOEHLSKEN-----ADLKAESEQ 1742

## RESULT 12

S28261

centromere protein E - human

N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E

C:Species: Homo sapiens (man)

C&gt;Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 19-Jan-2001

C:Accession: S28261

R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

Nature 359: 536-539, 1992

A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A:Reference number: S28261; MUID:93024922  
A:Accession: S28261  
A:Molecule type: mRNA  
A:Residues: 1-2663 <YEN>  
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865  
C:Genetics:  
A:Gene: GDB:CENPE  
A:Cross-references: GDB:361164; OMIM:117143 -  
A:Map position: 4q24-q25  
C:Superfamily: centromere protein E; kinesin motor domain homology  
C:Keywords: Atp; coiled coil; microtubule binding; mitosis; nucleotide binding; p-loop  
F:7-335/Domain: kinesin motor domain homology <KMT>  
F:8-93/Region: nucleotide-binding motif A (p-loop)  
F:486-2183/Domain: coiled coil #status predicted <COI>  
F:92/Binding site: Atp (lys) #status predicted

Query Match 3.28; Score 229; DB 1; Length 2663;  
Best Local Similarity 18.2%; Pred. No. 0.0059;  
Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;  
QY 17 PGVKEISDKISKEEMVRRLKMWKTFMDMDODSEE-----EKELYLNLAHLASD 66  
DB 333 PYVNEVS---TDEALLKRYR---KEIMDLKKOLEVSLSTRAQAMEKD---QLAQLLEEK 383  
QY 67 FFLHPGKDVRLLVACCLADIFRIYAPEPYTSPDKLKDIF-MFITRLQKGL-EDTSPQ 124  
DB 384 DLLQK-----VQNEKIENLRLMLVTSSSLTLQOELKAKR 417  
QY 125 FNRFFYLENTAWKSYNICEFEEDSNEIFTQLYRTLPFSVINNGHNGKVVHMMVDLMSI 184  
DB 418 KRRVTWCLGKINKMNSN-----YADOFNIPTN-ITTKTHKLSINLLREI 461  
QY 185 ---ICEGDTVSQELLDVFL-VNLVPAHKNLKNQAVDLAKALLKRTAQAIPIVITTFNQV 240  
DB 462 DESVCSDESDFSNLDTLSEIENWPATKLLNQENIE----- 497  
QY 241 LMLGKTSISDSEHFVFLILELYNIDSHLLSVLPQLFEFLKLSNDNEERLQVVKLLAKMF 300  
DB 498 -----SELNSLRADYDNLV-----DVEQLRTEKEEMELKLEKNDLDEFEALERKTK-- 545  
QY 301 GAKDSELASQNKPLWQCYLGRFNDIHPVIRLECYKFASHCLMNHDPDLAKDLTEYLKVESH 360  
DB 546 --KQEMQ-----LIHEISNLKLVKREYVNO 571  
QY 361 DPEAIRHDVIVSIVTAAKDLLVNDHLLNFVRERTLDKRWVRKKEAMGLAQIKKYA 420  
DB 572 DLENEL-----SSKVLELREKEDQIKKQIYIDSK--LENIKMDLS-----YS 613  
QY 421 LQSAAGKDAAKQIAWIKDKLHIYYQNSIDDRLLVERIFAQYVMVPHNLETTERMKCLYYL 480  
DB 614 LESTEDPKMQKOTLFDATV-----ALDAKRSAFLRSENLELEKEMKELATT 661  
QY 481 YATLDNALNALNEMWKNQNLRLHQVKDLDLIHKPTDASVKAFISKVMVITR----- 534  
DB 662 YKOME-----NDIQYQSOLEAKKMQVDLEK-----LQSAFNEITKLTLSDIGKV 708  
QY 535 -----NLDPGKAQDFMKKFTQVLEDDKIRKOLEVLVSPSCCKQAEQGVRETKKLG 588  
DB 709 PKDLLCNLEEGKTYDLQKELNKEVEENEALREEV-ILLS---ELKSLPSEVERLKEI- 763  
QY 589 NPKOPTNPFLEMIKFLLERIAPVHIDTES-ISALIKOVNKSIDGTADDEGVPTDQAI- 646  
DB 764 ---QDKSEELHIITSEKDKLFSVVVHKESRVQGLEEIGTKDKDLATTQSNKYKSTQOBFQ 820  
QY 647 -----RAGLELLKVLSTFHPISFSAETFESLACLKMD----- 680  
DB 821 NFKTLHMDFEOKYKMWLEENRMNQEIUNLSK-----BAQFDSLSLGAALKTSLSYKT 872  
QY 681 ---DEKVAEALQIFKNTGSKIEDFPHIRSALLPVLLHKS-----K 719  
DB 873 QELOKRETVQERL--NEMEQLKQLENRDSPLQTVREKTLITEKLOOTLEEVKTLTOE 930











47 VEODLSSSVOKALHPMRAVLSAD--LLRNPDSVVRVSVVSCUTEIMRITAPAPY-NDQ 104  
103 LKIDPFITRQIKGLDTPKSPNRYFYLLNIAWVKSNNICFELEDSNEITQYRILF 162  
105 MKDIFQVITAEFKLADASSRYSRAEVLFTLVAKVRSVLMDLE-CDLVLEMFQRL 163  
163 SVINNGHOKVHHMVLDLMSIIICGDTVSOELLDVLVNLVPAHKNLKNQAYDLAKALL 222  
164 KIIRPDHPLVLSMETIMITVIDESEVPMDLLELLTIVTKDSQDVSPAALTIVKVL 223  
223 KRTAQAEIYITFFNOVLMLGKTSISDLSEHVFDLILELYNIDSLLSLVLPQLEF--- 279  
224 SSTCKLQPCIM---EALKSSGTS-----LDMSY---PVVSSICQSEFAT 263  
280 -----KLKSNONEERL---QVY--KLLAKMGAKDSELASQNKPLWQCYLGRFNDIHP 328  
264 QAHNDVKPKDNEADEKISGEQVVPDSLEDKLNLGLSRKGRSARGGTRRANGD--- 320  
329 IRLCEYKFAASHCLMHPDLAKDLTEYLKVRSHDPPEAIRHVDIVSVITAAKKDILLVNDH 388  
321 ---EKVITANEGLESTD-AETASGSTRKRGWKPKSLMNPPEGYSFKTSSSKV----- 370  
389 LLNFVRETLDRWRVRKEMAMGLAQIYKKYALQSAAGKDAQAQIAWIKDKLLHIYQNS 448  
371 -----QEKELG-----DSSLGRVAAKKVP----- 389  
449 IDORLLVERIFAQYVWPHNLETTERMKCLYLYATLDLNAVNALNMWKCQNLLRHQVKD 508  
390 -----LPSRVGQTNQ-----SWISLSSSGRARTGSRKRSR- 420  
509 LLDLIKQPKTDASVKAIFKVMVITRNLPDPGKAQDFMKKFTQVLEDDK--IRKOLEVL 566  
421 ----TKMETDHDVSSVATQ-----PAKKQT-VKKTNPAREDLPKSNVKKHEDGI 465  
567 VSPTCCKQA--EGCVREITKKLGNPKQTPNPFLE--MIKFLLERTAPVHIDTESALI 622  
466 KTKCKSKKEKADNGLAKTSAKK-----PLAETMMVKPSGKKL--VHSDAK----- 508  
623 KQVNSIDGTADDEGVPTDQAIRAGLELLKVLSTPHSPHSATFSSLLACLKMDDE 682  
509 ---KKNSEGASMD---TPIPOSSKS-----KKKDS 532  
683 KYAEALQIFKNTGSKIEEDFPHIRALLPVLHKKSKGPPQAKYAIHCIIHAIFSSKET 742  
533 RATTATK-----KSEQAPKSHPKMKR-----IAGEEV 560  
743 QFAQIPEPLHKSLDPSNLEHLITPLVTIGHIALPADQFAAPKSWVATFIVKDLLMND 802  
561 E-----SNTNEL-----GEE 570  
803 LPQKTKTKLWVPDEVSPTMVKIQAIKMWVRWLLGMKNHNSKSGTSLRLTLILHSDG 862  
571 LVGRVNVWMLPDKFYEYGVKSYCRVKM-----HQ-----VTYSOG 608  
863 DLTEQCKISKPDMSRLRLAAGSAIVKLAQPCYHEIITLEQYQLCALINDECYQVRQVF 922  
609 DV-EELNLLK---ERKTIEDSSASEDDLESTPLSAF----- 646  
923 AOKLHGLSRLPLPLEYMATCALCAKDPVKERRAHARQCLVKNINVRREYLAKOHAASEK 982  
647 -----VEP-----SSSPEVRSSMQTKKKD-----IQREKSKRKIVSKN 661  
983 LLSLLPEYVVPYTHLLAHDPPVVKVQDTEQLKDVKECLWFLVLEILMAKNENNSHAFIRK 1042  
662 -----VEP-----SSSPEVRSSMQTKKKD-----S 682  
1043 MVENIKOTDAQDDAKKNKLYTVCDVAMNIIKSTYTSLESKDPVLPARFFTPQD 1102  
683 VTDISIKOTRTKALKAVSNP-----ESTGKNLKSUK-----KLNGEPD 723  
1103 KNFSNTKNVLPPEMKSGFFPGPKTTNVLGAVNKLSSAGKQSKTSRMTVSNASSSS 1162

Db 724 KTRGRT-----GKKQ-----KVTOAMHRKIEK----- 745  
QY 1163 NPSSGRIKGRLDSSMDHSENYTMSPLPKKSKDRDDSDLVRSLELEKPRGRKKTVP 1222  
Db 746 -----DCDEQEDLETKEDEDSUKL-GKESDA--EPDRMEDHOELPENHNVTYK 790  
QY 1223 TBOEEKLGMDLTKLVQEQPKGSQSRKRGHTASBESDQWPEEKRLKLEIDLE--NEDE 1280  
Db 791 TDGEE-----QEAAKEPTAESKTNGEENAEPTDCKEHSKLEPNAEPKSDGE 839  
QY 1281 QNSPPKK-----GKRGRPKPLGGTTPKEEPTMKTS-----KKGSKKKSGPPAPEEE 1328  
Db 840 EQEAAKEPNAELKTDGQENQAAKELTAERKTDDEEHKVADEVEQKSOQKTNVEPEAGEE 899  
QY 1329 EEROSGNTQOKSKQHRVSRRAQRAESPSSAIESTQSTPQKGRGPRSKTPS 1383  
Db 900 QKSVEEPNAEPKTK-----VEEKESAKEQTADTKLIEKEDMSKTGGEIDKETYS 949  
RESULT 9  
S38173  
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YKR095w; protein YKR415  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000  
C:Accession: S38173; S40647; S31207  
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale  
submitted to the Protein Sequence database, March 1994  
A:Reference number: S38158  
A:Accession: S38173  
A:Molecule type: DNA  
A:Residues: 1-1875 <BAL>  
A:Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w  
R:Experimental source: strain S288C  
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;  
Yeast 9, 1349-1354, 1993  
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr  
A:Reference number: S40644; MUID:94205265  
A:Accession: S40647  
A:Molecule type: DNA  
A:Residues: 1-1875 <BOU>  
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554  
R:Experimental source: strain S288C  
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.  
Mol. Gen. Genet. 237, 359-369, 1993  
A:Title: A new yeast gene with a myosin-like heptad repeat structure.  
A:Reference number: S31207; MUID:93247549  
A:Accession: S31207  
A:Molecule type: DNA  
A:Residues: 1-300, 'A', 302-1875 <KOE>  
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959  
C:Genetics:  
A:Gene: SGD:MLP1  
A:Cross-references: SGD:S0001803; MIPS:YKR095w  
A:Map position: 11R  
Query Match 3.3%; Score 235.5; DB 2: Length 1875;  
Best Local Similarity 19.5%; Pred. No. 0.0018;  
Matches 301; Conservative 246; Mismatches 602; Indels 391; Gaps 72;  
QY 19 VKEISDKISKE-----EMVRLKVVVKTfMDMDQDSEEEKEL-VLNLAHLASDFLKH 71  
Db 278 IGLSDSLSEKQEPFAEMSLKQRLVDLLESQJNAVKEELNSIRELNTAKVIADDSKKQT 337  
QY 72 PG-----KDVRLI---VACCLADIFRI-----YAPEAPYTSDDKLDKIDFMFITRLKGLLE 118  
Db 338 PENEDLLKELQTLKKEKLAQCEKLECLSLSTTDEADENLAKSSSSDFFLKLKLIKER 397  
QY 119 DTK---SPOFNRYFYLLN-IAWVSYNTCFELEDSENEIFTQYRITLFSVINNGHOKVH 174  
Db 398 RTKEHLQNTIETFIVELEHKVPILNSFK---ERTDMLNELNNAALLLETSTSEKNAKVK 454







C:Species: *Saccharomyces cerevisiae*  
C:Date: 08-Jul-1995 #sequence\_revision 09-Mar-1996 #text\_change 29-Oct-1999  
C:Accession: S54451; S52836  
R:Gentiles, S.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54451  
A:Accession: S54451  
A:Molecule type: DNA  
A:Residues: 1-682 <GEN>  
A:Cross-references: EMBL:249259; NID:g807956; PIDN:CAA89222.1; PID:g807957; MIPS:YMR0766  
A:Experimental source: strain AB972  
R:Pearson, D.; Bowman, S.  
submitted to the EMBL Data Library, April 1995  
A:Reference number: S52814  
A:Accession: S52836  
A:Molecule type: DNA  
A:Residues: 659-1277 <PEA>  
A:Cross-references: EMBL:248952; NID:g763008; PIDN:CAA88801.1; PID:g763023; MIPS:YMR0766  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:PDS5  
A:Cross-references: SGD:S0004681; MIPS:YMR0766  
A:Map position: 13R

Query Match 7.7%; Score 557; DB 2; Length 1277;  
Best Local Similarity 20.3%; Pred. No. 4.7e-19;  
Matches 272; Conservative 267; Mismatches 611; Indels 190; Gaps 46;

QX	12	KIYPPGVKEISDK - ISKEWVRRLKMWKVTFMDDODSEEEKEL ---YINLALHLASDDFF	68
DB	8	KLAFNSPIISTQSLISTNELLDRKLAAHELASLODQDNTDLTGLDKYRDA --- LVSRKL	64
QX	69	LKHPGKDVRLLVACCLADIPIRIYAPAPYTPSPDKLIDFIMFITROLKGLDFTKSPOFNRY	128
DB	65	LKKHGVIRAFACCLSDILRLVAPAPYDA -QLTDIFKLVLSEFQGLDQGVHIOQ	123
QX	129	PYLLENTAWKSYNICFEBSDSNEIFTQLYR-----TLFSVINNGHNQKVHMH	177
DB	124	TYLITKLELYRSTVLADLPSSNNLIELFHFIFDPNKSFPARLENVIGG-----	173
QX	178	VDMSSITCEGSDTVSQELDITVL-----VNLVPAHKLNKO-ADYDLAKALLKRTAQAI	229
DB	174	--ILGEVISEFSDVPUEVLRLIFENKFLTYPNPIPEGLNVTDCGVSESLICDTSYNSM	231
QX	230	EPYITTFNFVNLML-----GKTSISOLSEHVFDLIELYNIDSHLLSVLPQLEFKLKN	284
DB	232	SRHLTKYSEIIHEATNDDNNSRLTVVYKHLKVLRLWETPELINAVIGFYHELSE	291
QX	285	DNEERLOVVKLLAKMFGA -KDSELASONKPLWOCYIGRFENDIHVPIRLBCVXFASHCLN	343
DB	292	NELFREATEKLGIOILTSYSDLNFVSTHSTFKAWISKIADISPDVRVWETESIPQIIAT	351
QX	344	HPDLAKDLTVLKVRSHPDEEARHVDVIYSI---VTAAKKDL--LVNDHLLNFVRRT	397
DB	352	REDISKELNALAKTFIDSDPRVRSTVMIKNVPVTEIKWNITNKAITSLHLHARE--	409
QX	398	LDRKVRVRKEAMGLAQIYKKYALQ---SAAKGDAARQAWIKDKLHLHYIYQNSIDRLL	454
DB	410	--KHKEVRELINTMAKFYSNSLNEIERTYONKEIWEIIDTIPSTLYNLYINDLNIHQ	467
QX	455	VERIFAQYVWPHNLLETTERMKCLYLYLATLDLNAVAKALNEMKQCNLLRHQVKDLDLIK	514
DB	468	VDSVIFELLYPFPDNDKRVHRLTTLVLSHFDKKAFTSFFAFNAQRQTKISFAISKYIDFSK	527
QX	515	QPKTDASVKAIFSVMV-----ITRNLPPGKAQDFMKKFTQVLEDDDEKIRKOLEV	565
DB	528	FLNNOESMSSOGPIVMNKYNQTLWLASGLSDSTKAIDAEITIKOF--NDERIFYLNA	585
QX	566	LVSEPTSCQKQECVREITPKLGNPK-----OPNPFLEMTKFLLERIAPVH	612
DB	586	CVTNDIPFLTFKNCYNELYSKLOTGLFKKYNISTGASIMPRD-TAKVQIILLFRASPII	644

Qy	613	IDTESISALIKQVKNKSIDGTADDEGVPTDOAIRAGLE--LLKVLSTFTHPIST-HSAET	669
Dz	614	: :   :	:
Dz	645	YVNSNIVLNLNNNS-----DAQLDLKRRLTDDISKVNPTLFKDQIRT	689
Qy	670	FESLLACL-----KWDEKVAAEALQIFKNVGS-KIEEDPHIRTSALLPVLHHSKKGP	722
Dz	690	LKTIIDLODPDAEKNDNISLEALKTLTKASKTLKDQVDFO--TFFTFKLYDFAVESK	747
Qy	723	PRQAIAIHCHIAFSKETQPAQI--FEPLHKSLDPSLNELHLTPLVIGHIALAPD	779
Dz	748	PEITKYATKL-ALSAPAEETLKAKIRILPOLDKDYFTSHIILVMEIFEKKPPHVUND	806
Qy	780	QFAAPNKSWATFVKDLLMNDRLPKGKTTKL-WYPDEEVSPETWV----KIQAIKMVMR	834
Dz	807	D-----STDIIISYLIREVLLSNQVGDSKKEIDWDSDLSLDTKYSAIGNKVFCLKLPFN	861
Qy	835	WLGM-----KNHNSKGT-STLRLLTTTLHSDGDL-TEQGKISKPMDS----	888
Dz	862	KLRSIAPDVPRDELAESEFTKMTKFFYLIASGGELISEFNKEFYPTPSNYOTKLRCVAG	921
Qy	884	SATVKLAQECYHEITLBOYQLCALAINDECYQYQVFAQKLHGSLSRLLPLEYMAIC	943
Dz	922	IQVLKLARISNLANFIKPDSITKLINLVEDESLPVRKTFLEQLDKDYVANELISIKFLPV	981
Qy	944	ALCAKDPVKERRAHACQCLVKVINRVREYKLGHAAVSEKILLSLPEYVPVTIHLAHPD	1003
Dz	982	FFTAYEPDVELKTTKIWI--NFTGLCKSEKK-----GTFERALPLRIHAIAHP	1030
Qy	1004	DYKVVQDIE-----QLKDVKECLMFVLEILMAKNENSHAFIRKMVENIK-----QT	1051
Dz	1031	DIVGDLSEGDAYLNALTATDYLLFYFDISAQENFSLYYLSERVKNYQDKLVEIDE	1090
Qy	1052	DAQPDPAK-----MNEKLYTVCDVANNIWS-----KSTYS-----LES	1095
Dz	1091	EEEGPOKEAPKRRPYGOKMKYIGELSOMILLNLKKNQHMSAYPGKLNLPDLFKP-	1149
Qy	1096	RFTFPQDNFSNTKNYLPPEMKSFPTPGPKPTTNVLGAVNKPSSAGKQSOTKSRMETV	1155
Dz	1150	-FATVQEALIS-FKTYIPESLT-----EKIONNKAKIGRILHTSQTORLOQRLLAH	1201
Qy	1156	SNASSSNSPSPGRIKRGLSDSMDHSENEDYTNSPLPGKSD-KRDDSDOLVRSELEK	1214
Dz	1202	ENNESQKKKKKVHHARSQADDEEGDGDRSDSDSDSYSPSKNETKKGHENIV---MK	1258
Qy	1215	RGKKTPVTQEELGMDLL 1234	
Dz	1259	RVRKEVDYKDDDED---DBI 1274	

RESULT 6  
T43647  
hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C:Accession: T43647  
R:Lee, M.; Yoo, H.S.; Chung, K.S.  
submitted to the EMBL Data Library, February 1998  
A:Description: Clone 17 (bimD homologue).  
A:Reference number: Z22596  
A:Accession: T43647  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-390 <LEE>  
A:Cross-references: EMBL:AF049529; PID:AAD02493.1

Query Match 4.3%; Score 309.5; DB 2; Length 390;

Best Local Similarity 24.8%; Pred. No. 6.6e-08;  
Matches 105; Conservative 76; Mismatches 192;  
Indels 51; Gaps 11;

**QY** 825 KIOAIKMMVRLWLLGKNNHS--KSGTSTLRLLTLILHSDGDLTFEQKISKPDMSRLAA 882  
| : | : : : | | : : : | : : : : | : : : : | : : : : |  
**Dp** 4 KVLAIRVLNRLRAAGGTGAALNIGADPIKLKLVLMADGELSPFKNTPKISRVLRIITA 63















GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:00:54 ; Search time 26.15 seconds  
(without alignments)  
4051.962 Million cell updates/sec

Title: US-09-512-581-2  
Perfect score: 7193  
Sequence: 1 MAHSKRTNDGKITPPGVK.....QKGRGRPSKTPSPQPKNV 1391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
PCst-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2698.5	37.5	851	2 T00374	hypothetical prote
2	807	11.2	1579	2 T23142	hypothetical prote
3	632.5	8.8	1506	2 S22957	blind protein - Eme
4	571.5	7.9	1303	2 E96805	hypothetical prote
5	557	7.7	1277	2 S44451	hypothetical prote
6	309.5	4.3	390	2 T43647	hypothetical prote
7	303	4.2	780	2 F96840	hypothetical prote
8	264	3.7	990	2 H86293	protein T24D18.4 [
9	235.5	3.3	1875	2 S38173	myosin-like protei
10	234.5	3.3	852	2 T06310	hypothetical prote
11	231.5	3.2	2253	2 T30336	hypothetical prote
12	229	3.2	2663	1 S28261	nuclear/mitotic ap
13	227	3.2	1916	2 S46157	centromere protein
14	223.5	3.1	4717	2 T41581	RIF protein - yea
15	223	3.1	2139	2 T41896	hypothetical colle
16	222.5	3.1	1819	2 A71928	myosin heavy chain
17	222.5	3.1	2301	2 T02323	cag island protein
18	221.5	3.1	1927	2 G64585	nodulin-like prote
19	220.5	3.1	1727	2 T50073	cag pathogenicity
20	217.5	3.0	1020	1 QFHUH	myosin-like coiled
21	216.5	3.0	1790	2 S67593	neurofilament trip
22	210	2.9	1072	1 A37221	transport protein
23	209.5	2.9	1365	2 T45031	neurofilament trip
24	207.5	2.9	854	2 S02003	hypothetical prote
25	207.5	2.9	3488	2 T34418	neurofilament trip
26	206.5	2.9	699	2 T38073	hypothetical prote
27	206	2.9	6842	2 T29757	nucleolar phosphop
28	205	2.8	2094	2 S33124	protein UNC-89 - C
29	203.5	2.8	2364	2 A56577	tpv protein - huma
					microtubule-associ

ALIGNMENTS

RESULT 1

T00374

hypothetical protein KIAA0648 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00374

R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura,

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl

A:Reference number: Z14142; MUID:98403880

A:Accession: T00374

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-851 <ISH>

A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0648

Query Match 37.5%; Score 2698.5; DB 2; Length 851;  
Best Local Similarity 62.1%; Pred. NO. 1.1e-121;  
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;

Qy	477	LYLYATIDLNAV	KALNEMKCONLLRHQV	KDLDDLIKOPKTDASV	KAFSKVMVITRNL	536
Db	1	LYLYASLDPN	AVKALNEMKCONMLRSH	VELLDLHKQPTSEANG	CSAMFGKLTIAKNL	60
Qy	537	PDGKAQDP	PMKFTOVLEDEKIRK	QVLEVSPTSCQAE	GCVRITKKGPNKQPTNP	596
Db	61	PDGKAQDP	EVKFNQVLDDEKLRSQ	LELLISFTCSCKQAD	ICVRIAKRLANQPTNP	120
Qy	597	FLEMIFLERI	APVHIDFESISALIKOV	NSIDGTADDEGVPTDQ	AIKRAGLELLKVL	656
Db	121	FLEMVKFLER	APVHIDSEATISALV	KLMKNSIEGTADDEEG	VSPTAIRSLELLKVL	180
Qy	657	STHPSIFSH	SATFESLLACLKMD	DEKVAEALQIFKNTG	SKEEDFPHRSALLV	716
Db	181	STHTPTSF	HSATFESLLQCLRM	EDDKVAEAAIQFRN	TGHKTIETDLPQIR	STLPIIDHQ 240
Qy	717	KSKKGP	PPRQKVAIHCHIA	IFSCSTQFAQIFEP	HLKSLDPSNLEHLIT	PLVTIGHIALL 776
Db	241	KAKRGTPH	QAKVAHCIAHFTN	KEVQLAQIFEPLSR	SLNADVPVEQLIT	PLVSLGHISML 300
Qy	777	APDQFA	APKWSVATFIVK	DLMLNDRLPGKTK	TKLWVPDEEVSPT	MTVMKIQAIKMMVRWL 836
Db	301	APDQFASP	KMPKSVVNFIVK	DLMLNDRSTGEK	NGKLPDEEVSPE	VLAKVQAIKLVRL 360
Qy	837	LGKMNH	SKSGTSTLRLLT	ITILHSDGDLTEOG	KISKIPDKMSRL	RLAAGSAIVKLAQPCYH 896
Db	361	LGKMNQ	SKSANSUTRLLS	AMLVSEGLDETEQ	KRISKSDMSRL	RLAAGSAIMKLAQPCYH 420























## RESULT 11

```

NUM1_YEAST
ID NUM1_YEAST STANDARD; PRT; 2748 AA.
Q00402;
01-DEC-1992 (Rel. 24, Created)
01-DEC-1992 (Rel. 24, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEAR MIGRATION PROTEIN NUM1.
GN NUM1 OR YDR150W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE=92079907; PubMed=1745235;
RA Kormanec J., Schaaff-Gerstenschlaeger I., Zimmermann F.K.,
RA Perecko D., Kuentzel H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
RT highly repetitive 313 kDa NUM1 protein.";
RI Mol. Gen. Genet. 230:277-287(1991).
CC -!- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS
CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
CC ENVELOPE.
CC -!- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
CC FLANKING DOMAINS OF THE TANDEM REPEATS.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X61236; CAA43554.1; -.
CC PIR; S19052; S19052.
CC SGD; S0002557; NUM1.
CC InterPro; IPR001849; -.
CC Pfam; PF00169; PH; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC Repeat.
CC DOMAIN 593 1384 12.5 X TANDEM REPEATS.
CC FT REPEAT 593 656 1.
CC FT REPEAT 657 727 2.
CC FT REPEAT 728 798 3.
CC FT REPEAT 799 862 4.
CC FT REPEAT 863 926 5.
CC FT REPEAT 927 990 6.
CC FT REPEAT 991 1054 7.
CC FT REPEAT 1055 1118 8.
CC FT REPEAT 1119 1182 9.
CC FT REPEAT 1183 1246 10.
CC FT REPEAT 1247 1310 11.
CC FT REPEAT 1311 1374 12.
CC FT REPEAT 1375 1384 13 (INCOMPLETE).
CC DOMAIN 2573 2683 PH.
CC SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

```

Query Match 2.8%; Score 202; DB 1; Length 2748;

Best Local Similarity 18.2%; Pred. No. 0.11;

Matches 273; Conservative 229; Mismatches 583; Indels 416; Gaps 63;

23 SDKISKEVMYRLKVMYKTPMDQDSEEEKLYLNLALHLASDFFLKHPGKDVRLLVAC 82

```

Db 116 SSNEAKNEQIKSLQIKESLSDKIEELTNQKSGFMK-ELDSTKDLNWDLESKLTNLSMEC 174
QY 83 CLADIERIYAPEAPYTPDK- - - - -LKDIFMFTROLKGLDTSKPFQNY- - - - -FVLE 133
Db 175 --ROLKELAKKTEKSWNDEKESLKLKLTDLLEILTTLKNGMENDLSSQKLYHDEISELKE 232
QY 134 NIAWAKSYN- - - - -ICFELEDSNEIFTQLYRTLFVSINN 167
Db 233 RIILDNENDRLISVSLTSEINSLQSNTERIKIKQKLDLDAKASISSLKRK- - - - -VQK 288
QY 168 GHNQKVMHMMVDMSSIIICBGTVSQELLDTV- - - - -LVNLVPAHK 208
Db 289 KYYKQKHTSDTTVTSDPDSGGTTSEEDIFDIEIDHMIETGSPVEDISEDILVKKYSEKN 348
QY 209 NLNQAYDLAKALLKRTAQAIETPIYITFFNQVL- - - - -MLGKTSISDLSHVFDLILELY 263
Db 349 NMILLNSDSYKNLLQKSESASKPKDDELMTKEVAENLNMIALPNDNDNYSKKEFSL- - - - - 403
QY 264 NIDSHLL- - - - -LSVLPOLEFKLKSNDNE- - - - -ERLQVVKLL- - - - -AKMEGAK 303
Db 404 --ESHIKYLEASGYKVLPLEEFE- - - - -NLNESLSNPSYNYLKEKLOALKKPIQDSIFNLL 458
QY 304 DSELASONKPLW- - - - -QCYL- - - - -GRFNDIHVPIRLECVKFASHCLMNHDPDLAKDLTEYLKY 357
Db 459 KEPTIDFLPLTSKIDCLIIPTKDYNDLFESVKNPSTEQMKKCLEAKNDLQSNICKWLEE 518
QY 358 RSHDPEAIRHDVIVSIVT- - - - -AAKRDILLVNDHLL- - - - -NFVRER 396
Db 519 RNCG- - - - -KWLNSNDLYFSMVNKIETPSKOYLSDKAREYDQVLIDTKALEGLKNPTIDFLREK 576
QY 397 TLDKRWVRK- - - - -EAMGLAQIYKQYALQSAAGKDAK- - - - -QIAWIKD- - - - -K 439
Db 577 ASASDYLKLLKEDYVSPSLYLVVEHAKATNHHLLSDSAYEDLVKCKENPDMFLKESAK 636
QY 440 LHHYIONSIDRL- - - - -LVERIFAYQVWPHNLETTERM- - - - -KC- - - - -LYY 479
Db 637 LGHTVVSNEAYSELEKLEOPSLEYLVVEHAKATNHHLLSDSAYEDLVKCKENPDMFLKE 696
QY 480 LYATLDLNAV- - - - -KALNEMWKNLLRHQVKD- - - - -LL- - - - -DLIKQ 515
Db 697 KSAKLGHTVVSNEAYSELOQRYSELEKEVEQPSLAYLVVEHAKATDHHLLSDSAYEDLVK- 755
QY 516 PKTDASVKAIFSVMVITRNLPDPGKAQDPMKFTQVLEDEDEKIRKOLEVLSPTSCCKQ 575
Db 756 CKENPDVEFLKESAKLGHVTVSSEYSELQRYSELEKEVE- - - - -QPSLAYLVVEHAKATDH 813
QY 576 ---AEGCVREITKLGKPNKQPTNPFLEMIKFLERITAPVHIDTETISALIKQVKNKSDGT 632
Db 814 HLLSDSAYEELV- - - - -KCKENPDMFLKESAKLGHVTVSNEAYSELEKLEQ- - - - - 863
QY 633 ADDEDEGVPTDQAIRAGLELLKVLSTHPTISFH- - - - -SAETTESLLACLKMD- - - - -EKVA 685
Db 864 ---SLAYLVVEHAKATDHHLLSDSAYEDLVKCKENSDFEFLKESKA 905
QY 686 EAALQIFKNFG- - - - -SKIBEDFPHIRSALE- - - - -PVLHH- - - - -KSKKGP- - - - - 722
Db 906 KLGHVTVSNEAYSELEKLEQPSLAYLVVEHAKATDHHLLSDSAYEDLVKCKENPDMFLK 965
QY 723 PROAKYAIHCIIAIFSKETQFAQIFELPHKSLDPSNLEHLITPLVTYIGHIALLAPQOFA 782
Db 966 EKSALIG- - - - -HTVWSN- - - - -EAYSELEKLEQPSLEYLVVEHAKATNH- - - - - 1005
QY 783 APKSWVATFIVKDLLMNDRLPGKTKTLWVPDEEVSPETMVKIQAKMVRWLLGKNKN 842
Db 1006 ---HLLSDSAYEDLVKCKEN- - - - -PDMFLKESAKL- - - - -G 1036
QY 843 HSKSGTSTLRLLTILHSDGDLTEQGIKISKPDMSRLRLAAGSAIVKLAQPCVHEITILE 902
Db 1037 H- - - - -TVVSNEAYSELEKLEQPSLEYLVVEHAKATNHHLLSDSAYEELVKCK 1084
QY 903 Q- - - - -YOLCALAINDCEYQVQVFAOKLHGLSRLRLP-LEYMAICA- - - - - 944
Db 1085 ENPDVEFLKESAKLGHVTVSNEAY- - - - -SELEK- - - - -KLEQPSLEYLVVEHAKATNHH 1134

```











FT	DOMAIN	204	382	INTERACTS WITH RPA194.
FT	DOMAIN	384	587	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	DOMAIN	601	617	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	MOD_RES	563	563	PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT	VARSPLOC	241	241	K -> KVTITSVRAE (IN ISOFORM BETA).
FT	CONFLICT	3	3	D -> A (IN REF. 2).
FT	CONFLICT	133	133	R -> S (IN REF. 2).
FT	CONFLICT	291	292	YA -> SV (IN REF. 2).
FT	CONFLICT	456	456	S -> P (IN REF. 2).
SQ	SEQUENCE	699 AA;	73720 MW;	DFD4AD94EDF659FB CRC64;
Query Match 2.9%; Score 206.5; DB 1; Length 699;				
Best Local Similarity 22.3%; Pred. No. 0.011;				
Matches 93; Conservative 55; Mismatches 172; Indels 97; Gaps 13;				
OY	1045	ENIKOTKDAOGPDADAKMNEKLYTCDVAMNIMSKSTYSLESKDPVLPARFFTPQDPKN	1104	
DE	228	DDSEEEKAAATPKKTVPKQVAVKAPVKAATPTKRKSSSEDSSEDEEBQK	284	---KPMKN
OY	1105	FSNTKNYLPPEMKGFFTPGKPKTTNVLGAVNKPSSAGKQSQTKSSRMETVSNASSSN	1163	
DE	285	KPGVSYAPP-----PSAPPKKSILG-TQPPKAVEKQOPVSS--EDSSDESDDSSSE	334	
OY	1164	-----PSPGRIKGRBLSSEMDHSENYDTWSSPLPGKSKDRDSD	1205	
DE	335	BEKPPPTKAVYSKATKPPPKAKAESSSDSSDSE--DDEAPSKPAGTTKNSNKPAV	393	
OY	1206	LVRSELEKPRGKKTTPV-----TQEEKLGMDDLT-KLVOEQKPGKSORS	1249	
DE	394	TKRPAVKAAAPKQVGGGKLLTRKADSSSESESESESESESESESESESESESESESE	453	
OY	1250	RKRHTASEDEQWPEEKR---LKEDILENEDEQNSPPKGRGRPPKPLGGTTP----	1302	
DE	454	-----ALSLPAKQAPGSRSSSDSSDSSSESESESESESESESESESESESESESESE	507	
OY	1303	-----KEPTTKTSKSKKSGPPA-----PEEEE	1328	
DE	508	SAKKGAESSSSSSDDSSSE	567	
OY	1329	BEEROSGTEKSKSHRVSRRRAQRAESPESSESAIE-STOSTPOKGRGRPSKTPSP	1384	
DE	568	BEKKAAVVVSKSGSLKRRKQNEAKAEATPQAKIKLQTPNTPPKKGEKRASSP	624	
RESULT 9				
IO	GCG4_HUMAN	STANDARD;	PRT;	2230 AA.
AC	Q13439; Q14436; Q13270; Q13654;			
DT	01-OCT-2000 (Rel. 40, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY A 4 (TRANS-GOLGI P230) (256 KDA			
DE	GOLGIN) (GOLGIN-245) (72.1 PROTEIN).			
CN	GOLGA4			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RE	SEQUENCE FROM N.A.			
RA	MEDLINE=96215236; PubMed=8626529;			
RA	Erich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;			
RT	"Molecular characterization of trans-Golgi p230: a human peripheral			
RT	membrane protein encoded by a gene on chromosome 6p12-22 contains			
RT	extensive coiled-coil alpha-helical domains and a granin motif.";			
RT	J. Biol. Chem. 271:8328-8337(1996).			
RN	[2]			
RE	SEQUENCE FROM N.A.			
RA	Seelig H.P.;			
RL	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RE	SEQUENCE OF 131-2230 FROM N.A.			

RC	TISSUE=Placenta;		
RX	MEDLINE=96125112; PubMed=8537393;		
RA	Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;		
RT	"Molecular characterization of golgin-245, a novel Golgi complex		
RT	protein containing a granin signature.";		
RL	J. Biol. Chem. 270:31262-31268(1995).		
RN	[4]		
RN	SEQUENCE OF 524-672 FROM N.A.		
RC	TISSUE=Gastric fundus;		
RA	Balague C.;		
RL	Thesis (1994), Instituto municipal de investigacion medica, Spain.		
CC	-I- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-		
CC	GOLGI.		
CC	-I- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE		
CC	GOLGI MEMBRANE.		
CC	-I- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE		
CC	SPLICING.		
CC	-I- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND		
CC	IN HEPATITIS B.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; U41740; AAC50434.1; -		
DR	EMBL; X82834; CAA58041.1; -		
DR	EMBL; U31906; AAC51791.1; -		
DR	EMBL; X76942; CAA54261.1; -		
DR	MIM; 602509; -		
DR	InterPro: IPR000237; -		
DR	Pfam; PF01465; GRIP; 1.		
KW	Golgi stack; Antigen; Coiled		
FT	DOMAIN 133 237		
FT	COILED COIL (POTENTIAL).		
FT	DOMAIN 276 1011		
FT	COILED COIL (POTENTIAL).		
FT	DOMAIN 1033 1214		
FT	COILED COIL (POTENTIAL).		
FT	DOMAIN 1259 2152		
FT	VARSPLOC 2154 2185		
FT	TPYKGNLYHPDVSFLGPTFEYLRKLVFEY -> HLTKV		
FT	AICTIRSHSLENPLNLSICEKFLSI (IN ISOFORM		
FT	2).		
FT	MISSING (IN ISOFORM 2).		
FT	VARSPLOC 2186 2230		
FT	MISSING (IN ISOFORM 3).		
FT	VARSPLOC 2103 2109		
FT	FTSPRSIGF -> SWLRSSS (IN ISOFORM 4).		
FT	R -> K (IN REF. 3).		
FT	Y -> H (IN REF. 3).		
FT	T -> A (IN REF. 3).		
FT	K -> E (IN REF. 3).		
FT	T -> A (IN REF. 3).		
FT	K -> E (IN REF. 3).		
FT	K -> N (IN REF. 3).		
SQ	SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;		
Query Match 2.9%; Score 205.5; DB 1; Length 2230;			
Best Local Similarity 18.5%; Pred. No. 0.057;			
Matches 302; Conservative 256; Mismatches 559; Indels 517; Gaps 74;			
OY	20	KEISDKISKKEWVRLLKMWVTFMDMDQDSEEEKLYNLALHLASDFFL-----KHPGK	74
DB	529	KEQESLAELEELQKKAILESENKLRDLOQEAETRTILELESSLEKSKQENKNQSK	588
OY	75	DVRLLVACCLADIFRIYAPEAPYTPDK---LKDIFMTITROLKGLDTPKSPQFNRYFYL	131
DB	589	DL-----AVHLEAKKNHKEITVAVVEKHKTELESKHHQ-----	623
OY	132	LENIAWKSYNIC-----FELEDSNEIFTQLYRTLF---SVINNGHNOKVHMHVMDLMS	183
DB	624	--DALWTEKLQVLKQYQYTEMELREKCEQEKETLLKDKKEIIFOAHIEEMNEKTEKLDV	681
OY	184	ICEGDTVSOELLDTV-----LVNLVPAHKNLNKQAYDLA---	218















































Query Match	3.2%	Score 133.2	DB 22	Length 936
Best Local Similarity	0.8%	Pred. No. 4.9e-23		
Matches	6	Conservative 496	Mismatches 284	Indels 0
Gaps	0			
71	ataaaatatacgaagagagagatggtgagacgattaaagatggtctgtaaaactttttcgg	130		
786	ww	727		
131	atagcaccagagactctgaagaagaagaagagctttatttaaacctagctttacatcttg	190		
726	ww	667		
191	cttcagattttttctcaagcatcctggttaaagatgttcgcttactgtagcctgctgcc	250		
666	ww	607		
251	ttgctgatatctcagagattatgctcctgaagcctcttacacatccctcgataaactaa	310		
606	ww	547		
311	aggatattattgtttataacaagacagttgaagggctagaggatcacaaagagccac	370		
546	ww	487		
371	aattcaataggtatttttatttacttgagaacattgctgggtcaagtcatataacatat	430		
486	ww	427		
431	gcttgagttagaagatagcaatgaatttcacccagctatcacagaaccttatttcag	490		
426	ww	367		
491	ttataaacaatggccacaatcacagaagtcacatatgcacatggtagacctatgagctta	550		
366	ww	307		
551	ttatttgtgaagtgatcacagtgctcaggagcttttgatacgggttttagtaaatctgg	610		
306	ww	247		
611	tacctgctcataagaatttaacaagaagaatgatttggaagagctttactgaaga	670		
246	ww	187		
671	ggacagctcaagctattggccatatattaccacttttttaatacaggttctgattctg	730		
186	ww	127		
731	ggaaacatcatcacagatttgcagagcatgctttgacttaatttggagctctaca	790		
126	ww	67		
791	attattgatgctattgctgctctctgttttaccocccagcttgaattaaattaaagagca	850		
66	ww	7		
851	atgata 856			
6	wwwwww 1			

RESULT 14

AAF58262/c

ID AAF58262 standard; DNA; 936 BP.

AC AAF58262;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D2007.

DE

XX















CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SQ Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

Query Match 7.0%; Score 292; DB 21; Length 295;  
Best Local Similarity 100.0%; Pred. No. 2.9e-62;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QX 2497 gttcgatgctacttggaataaataatcacagtaaatcaggaaacttcccttaaga 2556  
DB 1 gttcgatgctacttggaataaataatcacagtaaatcaggaaacttcccttaaga 60  
QX 2557 ttgttaacaaatattgcatagtgatgagacttgacagacaggaggaaattagtaaa 2616  
DB 61 ttgttaacaaatattgcatagtgatgagacttgacagacaggaggaaattagtaaa 120  
QX 2617 ccagatattgcatgctgagacttgctgctggagtgctattgtgaagctggcacaagaa 2676  
DB 121 ccagatattgcatgctgagacttgctgctggagtgctattgtgaagctggcacaagaa 180  
QX 2677 cccgtttaccatgaaatcatcattagacaataatcagctatgtgcatgtatcatcaac 2736  
DB 181 cccgtttaccatgaaatcatcattagacaataatcagctatgtgcatgtatcatcaac 240  
QX 2737 gatgaatgctatcaagtaagacaagtgtttcccgagaaacttcacaaagcc 2788  
DB 241 gatgaatgctatcaagtaagacaagtgtttcccgagaaacttcacaaagcc 292

RESULT 8  
ID AAF18338 standard; DNA; 1217 BP.  
XX AAF18338;  
AC AAF18338;  
DT 14-MAR-2001 (first entry)  
DE Lung cancer associated polynucleotide sequence SEQ ID 357.  
XX Human; lung cancer associated protein; neuroprotective; cytotstatic;  
KW cardioactive; immunomodulatory; muscular active; vulnerary;  
KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200055180-A2.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05918.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX  
XX Ruben SM;  
XX  
XX WPI; 2000-587514/55.  
XX P-PSDB; AAB58462.  
XX  
XX Lung cancer associated gene sequences, referred to as lung cancer  
XX antigens, useful for treatment, prevention, and diagnosis of disorders  
XX such as lung cancer -  
XX  
XX Claim 1; Page 815; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cytotstatic; cardioactive;  
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
CC general; nephrotropic; antiinfective; gynecological; or antibacterial  
CC activity. The invention also includes antibodies specific for the  
CC protein or polynucleotide sequences. The lung cancer associated  
CC polynucleotide sequences may be used for detection of lung cancer,  
CC chromosome identification, as chromosome markers, and for numerous other  
CC diagnostic or research purposes. The proteins may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders. The proteins may also be used in the treatment of wounds and  
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
CC peptide AAB58549 are used in the course of the invention for the  
CC identification and characterisation of the polynucleotide and protein  
CC sequences.  
SQ Sequence 1217 BP; 445 A; 232 C; 270 G; 263 T; 7 other;

Query Match 6.1%; Score 253.8; DB 21; Length 1217;  
Best Local Similarity 62.2%; Pred. No. 1.3e-52;  
Matches 399; Conservative 0; Mismatches 242; Indels 0; Gaps 0;  
QY 2893 aatataaattgaaggcgagtagtatctgaagcagcagcagctgttagtgaataattatg 2952  
DB 2 aatataaattgaaggcgagtagtatctgaagcagcagcagctgttagtgaataattatg 61  
QY 2953 tctcttaccagagtagtattgtttccatatacaattcaccttttggcacatgaccagat 3012  
DB 62 tcaactgttgcctgaatattgtattccatattcattcattcattcattcattcattcatt 121  
QY 3013 tatgtcaaatgacagatattgaacaaacttaagattgaagaaatgtttgtttgtttgtt 3072  
DB 122 tttaacagatcacaaagattgtgatcagctctgtgatcatcaaaagattgtctgttcattg 181  
QY 3073 ctggaaattattatggctaaataaataaacagtcacgctcttcttctacagaaatgggta 3132  
DB 182 ctgtaagttttatgacaaagaatgaaacaaatagccatgctcttattgaagaatggca 241  
QY 3133 gaaatattaaacaaacaaagaatgcccaagcagcagcagcagcagcagcagcagcagc 3192  
DB 242 gagaacatcaagttaaccagagatgccagctctccagatgaatcccaagcagcagcagc 301  
QY 3193 ctgtacactgt 3252  
DB 302 ctgtacacagtagtattgt 361  
QY 3253 ttggaaattctctaaagacccggtactaccagctctgtttcttctacacacacacacacac 3312  
DB 362 gcagattccacaaagacccagctctcccaatgaaattttttacacacacacacacacac 421  
QY 3313 ttcaagtaac 3372  
DB 422 ttctgttaacgataagagttatattttcagaagacagagtagtacttctgttaacaggaag 481  
QY 3373 cctaaac 3432  
DB 482 ccaagcctctgt 541  
QY 3433 tctcagac 3492  
DB 542 cccatgtttagaagcactgtgcactgagactggaagcacaattatgaattgaattgaagc 601  
QY 3493 agctctctcgtgaagataaaggagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3533  
DB 602 aacccttcaacggaatcgatcaagggaacagaggttcaga 642

RESULT 9











602 tcaatagaggggacagcagatgatgaagaggggtgtaagtccagatcacagctatccgt 661  
1942 gcaggctctgaactgcttaaggtactctcaattacacatcccctcattctcattctgct 2001  
662 tcaggacttgaaactcttaaggtctgctcttttacacatccctcctgctccactctgca 721  
2002 gaacatttgaaatcattactgggtctgctgaataatggatgaaagaatgacagaagct 2061  
722 gagacatatgctcctgttacagtgcctaaagaatggagatgacaaggtgacagaagct 781  
2062 gcaactcaaaatttcaaaaacacaggaagcaaaatgaagagattttccacacatcaga 2121  
782 gctattcaaaatttgaataacaggtcacaataatgaacacagaccttcccagatcaga 841  
2122 tcagccttgctcctgctttacatcacaaatctaaaaagagcccccccgctcagccaaa 2181  
842 tcgaccttaattcccaattttacatacaaaaagcaaaaggagggttactccacaccaagcaaaa 901  
2182 tatgcatcatttgatccatcgatgataattttctagtaaaagagaccagtttgcacagata 2241  
902 caggctgtgcactgtatacacgcatattcaacaataaagaagctccagcttgcacagatt 961  
2242 ttggacccctgcataagagcctagatccaaagcaaacctggaaacatctcataaacaccattg 2301  
962 ttggccacctcagtgaggagctggaatgctgagtgcagaaacaacttataactccatta 1021  
2302 gttactattggtcatattgctccttgcacctgatacaatttctgctccttggaaatct 2361  
1022 gttctattggccacatttctatgtagcaacagatcagtttcttcccacaaatgaaatct 1081  
2362 tgggtgactcttcaatttgaagaatcttctcatgaatgatcggttccaggggaaaaag 2421  
1082 gtatgacaaattttatttgaagaatctgctaataatgacaggtcacaaggtgaaaaag 1141  
2422 acaactaaacttgggttccagatgaagaagatctctctgagacaatggtcacaattcag 2481  
1142 aatggaaaacttggtctccagatgaaggttctccctgaaatgactagcaaaaggtacag 1201  
2482 gctattaaaatgatggttcgacttggtacttggatgaaataatcacagtaaatcagga 2541  
1202 gcaattaaactctgtaaggtggctgtggtgatgaaacaacacagctcctaactctgcc 1261  
2542 actctacattgaagtgttaacaacaataatgtcatagtgatggagacttgacagaaag 2601  
1262 aattcaacctctggttattatcagagatgttggttagtgagggtagacctgacagagcaa 1321  
2602 ggaaaaattagtaaacacagatgtcagctgcagactgtgagactgtctgggagtgctattgtg 2661  
1322 aagaggatcagtaaatctgatagtctcgtctgogattagctgctggtagtgccataatg 1381  
2662 aagctggcacaagaacctgttaccatgaaatcatcacattagaaacaataatcagctatgt 2721  
1382 aagcttgctcagaacctgttaccatgaaatattaccaccaagaacagtttcaagctgtg 1441  
2722 gcaattagctatcaacgatgaatgctatcaagtaagacaagtggttcccacagaaccttcac 2781  
1442 gcacttgatttaataatgatgaggttaccacagtaaggcagatatttgcacagaagctgcat 1501  
2782 aaaggcccttccggtttacaggttccacttgatgatatatggcaatctgtgccccttggca 2841  
1502 aagggcacttgtaagttactgcctccattggagttatggcagatcttggcttggctgcc 1561  
2842 aaagatcctgtaaggagagaagagctcatgctaggcaatgttggtaaaaaataataat 2901  
1562 aaagatcctgtgaaggagaagaagacacagcagcaaatgtttactgaaaaatatcagt 1621  
2902 gtaaggcgtgagtatctgaagcagcagctgagctgttagtgaaaaaattattgtctctcta 2961  
1622 atacgcaggaatacacattaagcagaatccctatggtcactgagaaattattatcactgttg 1681  
2962 ccagagtatgtttccatatacaattcacccttggcagatgacccagattatgtcaaaa 3021  
1682 cctgaatatgttccatatacattgattccactgtagcccatgacccagattttacaaga 1741

QY 3022 gtacaggtattgaacaacttaagaatgtttaagaatgtcttgggtttgttctggaata 3081  
DB 1742 tcacaagatgtgatcagcttctgatatacaagaagtgctcattgttcatgctgaagtt 1801  
QY 3082 ttaatggctaaaaatgaaaaataacagtcacgctctttatcatcagaagaatggtgagaaatatt 3141  
DB 1802 ttaatgacaaaagatgaaaaacaatgaccttcttatgagaagaatggcagagaacatc 1861  
QY 3142 aaacaacaaaagatcccaagaccagatgatgcaaaaatgaatgaacacactgtacact 3201  
DB 1862 aagttacacagagatcccagctccagatgaaatccaaagacaaatgaaaacactgtataca 1921  
QY 3202 gtgtgtgatttgccatgataatcatcatgtcacaagagtagtactacatcacagtttggaaatct 3261  
DB 1922 gtatgtgtgtgctctctgtgtataaatagtataaagtgcttgtgcaatgcagattca 1981  
QY 3262 cctaaagaccgggtactaccagctcgtttcttcaactcaactcagacaagaatttcaagtaac 3321  
DB 1982 ccaagagaccagctcctcccaatgaaatgtttttacacaaactgaaaaggacttctgtaac 2041  
QY 3322 accaaaattatctgctcctgaaatgaaatcatatttttactcctggaacacttaaaaca 3381  
DB 2042 gataagatttatatttcagaaagacaaagtagtcttctgtaacaggaagacaaagcct 2101  
QY 3382 accaatgttctagagctgttaacaagccactttcatcagcaggaagcaactctcagacc 3441  
DB 2102 gctggagtagtagtgtagtaataaagcctttatcagcaaggaagaaacccatgtt 2161  
QY 3442 aaatcatcagaatggaacctgtaagcaatgcaagcagcagctcaaatccaagctctcct 3501  
DB 2162 agaagcactggcactgagactggaagcaatattaatgaaattcagagctgaacccctta 2221  
QY 3502 ggaagataaaagggagcttgatgttctgaaatgtagtccacagtgaaaatgaaga 3557  
DB 2222 accygaatctgcataaggggacagaggttcagagcgacgaacactggagttagtga 2277

## RESULT 4

AAV87629  
ID AAV87629 standard; cDNA; 439 BP.

XX AAV87629;

XX 12-FEB-1999 (first entry)

XX EST clone DY17.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;  
chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

XX W09845437-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-0506956.

XX 10-APR-1997; 97US-0837312.

XX (GEM) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070078/06.

XX New polynucleotides encoding human secreted proteins - derived from  
e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
ovary, pituitary, retina and colon cDNA libraries







	2010	c t t a t t a a c a g t g a a c a a a t c a a t a g a t g y g a a c a g c a g a t g a t g a a d a t g a g g t g t t	2069
	1921	c c a a c t g a t c a a g c c a t c a g a g c a g t c t t g a a c t g t t a a g g t a c t c t c a t t a c a c a t	1980
	2070	c c a a c t g a t c a a g c c a t c a g a g c a g t o t t g a a c t g t t a a g g t a c t o t c a t t t a c a c a t	2129
	1981	c c a t c t c a t t t c a t t c t g t c t g a a a c a t t t g a a t c a t a t t a c t g g t g t g t c t g a a a t t g g a t	2040
	2130	c c a t c t c a t t t c t g t c t g a a a c a t t g a a t c a t t a c t g g t g t g t g a a a t t g g a t	2189
	2041	g a t g a a a a a g t a g c a a a g c t g c a c t a c a a a t t t c a a a a c a c a g a a g a a a a t t g a a	2100
	2190	g a t g a a a a g t a g a a g c t g c a c t a c a a a t t t c a a a a c a c a g a a g a a a t t g a a	2249
	2101	g a g a t t t t c c a c a c a t c a g a t c a g c t g t c t c t g t t t a c a t c a a a a t c t a a a a a	2160
	2250	g a g a t t t t c c a c a c a t c a g a t c a g c t g t c t c t g t t t a c a t c a c a a a t c t a a a a a	2309
	2161	g g a c c c c c c g t c a a g c c a a a t a t g c a t t c a t t g t a t c c a t g c g a t a t t t t c t a g a a	2220
	2310	g g a c c c c c c g t c a a g c c a a a t g c a t t c a t t g a t c a t g c g a t a t t t t c t a g a a	2369
	2221	g a g a c c c a t t t g c a c a g a t a t t t g a g c t c t g c a t a a g a g c c t a g a t c a a g a c a c c t g	2280
	2370	g a g a c c a g t t t g c a c a g a t a t t t g a g c t c t g c a t a a g a g c e t a g a t c a a g a c a c c t g	2429
	2281	g a a c a t c t a a c a c c a t t g t t a c a t t g g t c a t a t t g c t c t c t c t c c t g c a c c t g a t c a a	2340
	2430	g a a c a t c t a a c a c c a t t g g t t a c a t t t g g l c a t t g c t c t c c t g c a c c t g a t c a a	2489
	2341	t t t g c t g c c t t g g a a a c t t g g g t a g c t a c t t t c a t t g t g a a g a t c t t c t c a t g a a t	2400
	2490	t t t g c t g c c t t g g a a a c t t g g g t a g c t a c t t c a t t g t g a a g a t c t t c t c a g a a t	2549
	2401	g a t c g g c t t c a g g g a a a a g a c a a c a a a c t t t g g g t t c c a g a t g a a g a g t a t c t c t	2460
	2550	g a t c g g c t t c a g g g a a a a g a c a a c a a a c t t t g g g t t c c a g a t g a a g a g t a t c t c t	2609
	2461	g a g a c a a t g g t c a a a a t c a g g c t a t t a a a t g a t g g t t c g a t g g c t a c t t g a a a t g a a a	2520
	2610	g a g a c a a t g g t c a a a a t c a g g c t a t t a a a t g a t g g t t c g a t g g g c t a c t t g g a a t g a a a	2569
	2521	a a t a a t c a c a g t a a a t c a g a a c c t t a c c t a a g a t t g t c a a c a a c a a t a t t g c a t a g t	2580
	2670	a a t a a t c a c a g t a a a t c a g a a c t t a c c t a a g a t t g t c a a c a a c a a t a t t g c a t a g t	2729
	2581	g a t g a g a c t t g a c a a a c a g g g g a a a a t a g t a a a c c a g a t a t g t c a g c t g a g a c t t	2640
	2730	g a t g g a g a c t t g a c a a a c a g g g g a a a a t t a g t a a a c c a g a t a t g t c a g c t g a g a c t t	2789
	2641	g c t g c t g g g a g t g c t a t t g t g a a g c t g g c a c a a g a c c c t g t a c c a t g a a a t c a t c a a	2700
	2790	g c t g c t g g g a g t g c t a t t g t g a g g c t g g c a a g a a c c c t g t t a c c a t g a a a t c a t c a a	2849
	2701	t t a g a a c a a t a c a g c t a t g t c a t t a g c t a t c a a c g a t g a a t g c t a t c a a g t a a g a c a a	2760
	2850	t t a g a a c a a t a c a g c t a t g t c a t t a g c t a t c a a c g a t g a a t g c t a t c a a g t a a g a c a a	2909
	2761	g f g t t t g c c a g a a a t t c a a a g g c t t t c c g t t t a a g c g t t c c a c t t g a g t a t a t g	2820
	2910	g f g t t t g c c a g a a a c t t c a a a g g c t t t c c g t t t a a g c g t t c a c t t g a g t a t a t g	2969
	2821	g c a a c t g t g c c t t t g c a a a g a a c t c t g t a a a g a g a a g a g a g a g c t a t g t a a g c a a	2880
	2970	g c a a c t g t g c c t t t g t c a a a a g a t c c t g t t a a g a g a g a g a g a g c t a t g t a g c a a	3029
	2881	t g t t t g g t g a a a a t a t a a a t g t a a g c g g g a g a t a t c t g a a g c a g a t c a g c t g t t a g t	2940
	3030	t g t t t g g t g a a a a t a t a a a t g t a a g c g g g a g a t a t c t g a a g c a g a t c a g c t g t t a g t	3089
	2941	g a a a a a t a t t g t c t c t c a c a g a g a t a t g t t g t t c c a t a t a c a a t a c a c t t t t g g c a	3000
	3090	g a a a a a t a t t g t c t c t c a c a g a g a t a t g t t g t t c c a t a c a a t a c a c t t t t g g c a	3149

QY	3001	catgaccacgattatgtc	aaagtacaggatattga	caaaccttaaa	gatgttaaga	aatgtg	3069										
Db	3150	catgaccacgattatgtc	aaagtacaggatattga	caaaccttaaa	gatgttaaga	aatgtg	3209										
QY	3061	ctttggtttgttctg	aaataattaa	tgctcaaaa	atgaaaa	taacagtcacgctttt	atc	3120									
Db	3210	ctttggtttgttctg	aaataattaa	tgctcaaaa	atgaaaa	taacagtcacgctttt	atc	3269									
QY	3121	agaaagatggtg	aaaaattataa	caacaaaaa	agatgcca	aaagacagatgata	caaaa	3180									
Db	3270	agaaagatggtg	aaaaattataa	caacaaaaa	agatgcca	aaagacagatgata	caaaa	3329									
QY	3181	atgaaatgaaaaa	ctgacactgtgt	gtgatgttg	ccatgaata	catcatgtca	aaagagt	3240									
Db	3330	atgaaatgaaaaa	ctgacactgtgt	gtgatgttg	ccatgaata	catcatgtca	aaagagt	3389									
QY	3241	actacatacagttt	ggaattctct	aaagacc	cggtactac	cagctcgtttctt	caactcaa	3300									
Db	3390	actacatacagttt	ggaattctct	aaagacc	cggtactac	cagctcgtttctt	caactcaa	3449									
QY	3301	ctgtgacaagaatt	tcagtaacac	ccaaaaattat	ctgcctc	ctcgtgaaat	gaaatcat	tttttc	3360								
Db	3450	ctgtgacaagaatt	tcagtaacac	ccaaaaattat	ctgcctc	ctcgtgaaat	gaaatcat	tttttc	3509								
QY	3361	actcttggaaaa	acttaaa	caaccaa	atgttct	gaggagctgt	taacaag	ccactttt	catca	3420							
Db	3510	actcttggaaaa	acttaaa	caaccaa	atgttct	gaggagctgt	taacaag	ccactttt	catca	3569							
QY	3421	gcaggcaagcaat	ctcagacc	aaatcatc	acgaat	tggaact	gaaact	taagcaat	gcaagc	agc	3480						
Db	3570	gcaggcaagcaat	ctcagacc	aaatcatc	acgaat	tggaact	gaaact	taagcaat	gcaagc	agc	3629						
QY	3481	agctcaaa	ttccaa	gtctc	ctgga	aaataa	aaagg	gaggttg	atagtt	ctgaaa	tgat	3540					
Db	3630	agctcaaa	ttccaa	gtctc	ctgga	aaataa	aaagg	gaggttg	atagtt	ctgaaa	tgat	3689					
QY	3541	cacagt	gaaaa	atgaa	gat	acacaa	tgctt	ctccct	ttgc	ggg	gaaaaa	agtgaca	ag	3600			
Db	3690	cacagt	gaaaa	atgaa	gat	acacaa	tgctt	ctccct	ttgc	ggg	gaaaaa	agtgaca	ag	3749			
QY	3601	agagac	actct	gattct	gtgaag	tctgaa	ttgg	agaag	cctc	agag	cagg	gaaaaa	aaac	g	3660		
Db	3750	agagac	actct	gattct	gtgaag	tctgaa	ttgg	agaag	cctc	agag	cagg	gaaaaa	aaac	g	3809		
QY	3661	cccg	tcac	agaac	agag	gagaaa	at	laggtat	ggat	tgact	tgact	aaagt	tggt	ac	agg	3720	
Db	3810	cccg	tcac	agaac	agag	gagaaa	at	laggtat	ggat	tgact	tgact	aaagt	tggt	ac	agg	3869	
QY	3721	cagaa	acct	taagg	cagtc	aggaat	tcgg	aaaaag	ggcc	atac	gcgctt	cagaat	ct	g	3780		
Db	3870	cagaa	acct	taagg	cagtc	aggaat	tcgg	aaaaag	ggcc	atac	gcgctt	cagaat	ct	g	3929		
QY	3781	gaac	ac	gagtc	ggagaaa	agggc	caaa	gagata	tata	ta	gaaaa	at	gaag	at	g	3840	
Db	3930	gaac	ac	gagtc	ggagaaa	agggc	caaa	gagata	tata	ta	gaaaa	at	gaag	at	g	3989	
QY	3841	cagaa	tagt	ccg	ccaaaa	aggt	taaa	gagcc	gac	ccaaaa	acct	ct	tggt	gg	aggt	3900	
Db	3990	cagaa	tagt	ccg	ccaaaa	aggt	taaa	gagcc	gac	ccaaaa	acct	ct	tggt	gg	aggt	4049	
QY	3901	acac	aaaa	gaag	agcc	aaat	gaaa	actt	ct	aaaa	agga	agcc	aaaa	aaat	ct	gga	3960
Db	4050	acac	aaaa	gaag	agcc	aaat	gaaa	actt	ct	aaaa	agga	agcc	aaaa	aaat	ct	gga	4109
QY	3961	ctt	ccag	cc	agag	gagga	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	4020
Db	4110	ctt	ccag	cc	agag	gagga	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	4169
QY	4021	tc	aaa	g	caaa	ag	caac	ag	tc	g	tc	aa	gg	ag	ag	ag	4080
Db	4170	tc	aaa	g	caaa	ag	caac	ag	tc	g	tc	aa	gg	ag	ag	ag	4229



CC cell transfected with the expression vector, and a method for producing  
 CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has  
 CC cytosolic activity, and acts to suppress cell proliferation. The AS3  
 CC gene is useful as a marker for the efficient diagnosis and treatment of  
 CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of  
 CC antisense agents for sequence specific modulation of gene expression. The  
 CC AS3 protein may be used in the treatment of disorders caused by aberrant  
 CC modification or mutation of a gene encoding an AS3 protein, misregulation  
 CC of the AS3 gene or aberrant post-translational modification of the AS3  
 CC protein. This sequence represents the human AS3 cDNA sequence with an  
 CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when  
 CC compared with the claimed AS3 cDNA sequence AAA28051.  
 XX  
 SQ Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 4173; DB 21; Length 5355;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctcattcaaaactagaccatgatggaaaaattacatatccgctgggtcaag 60  
 Db 150 atggctcattcaaaactagaccatgatggaaaaattacatatccgctgggtcaag 209

QY 61 gaatatcagataaaatatctaaagaggagatggtagacgattaaagatggttgaaa 120  
 Db 210 gaatatcagataaaatatctaaagaggagatggtagacgattaaagatggttgaaa 269

QY 121 acttttattgatatggaccagactctgaagagaaaggagctttatttaacctagct 180  
 Db 270 acttttattgatatggaccagactctgaagagaaaggagctttatttaacctagct 329

QY 181 ttacatcttggctcagatgtttttctcaagcatcctggttaaagatgttcgcttactggta 240  
 Db 330 ttacatcttggctcagatgtttttctcaagcatcctggttaaagatgttcgcttactggta 389

QY 241 gctctgctctgctgatatatttcaggatttatgctctgaagctccttaacatccct 300  
 Db 390 gctctgctctgctgatatatttcaggatttatgctctgaagctccttaacatccct 449

QY 301 gataactaaaggatattttattgtttaaacaagacagttgaagggtctagagatata 360  
 Db 450 gataactaaaggatattttattgtttaaacaagacagttgaagggtctagagatata 509

QY 361 aagagcccaaatccaataggtattttatttacttgagaacatgcttgggtcaagttca 420  
 Db 510 aagagcccaaatccaataggtattttatttacttgagaacatgcttgggtcaagttca 569

QY 421 tataacatatgcttggatgataagatgcaatgaaatttccaccagctatacagaacc 480  
 Db 570 tataacatatgcttggatgataagatgcaatgaaatttccaccagctatacagaacc 629

QY 481 ttattttcagttataaacaatggccacaatcagaagtcctatgcacatgtagacctt 540  
 Db 630 ttattttcagttataaacaatggccacaatcagaagtcctatgcacatgtagacctt 689

QY 541 atgagctctattatttgaaagtgatacagtgctctcaggagcttttggatcaggtttta 600  
 Db 690 atgagctctattatttgaaagtgatacagtgctctcaggagcttttggatcaggtttta 749

QY 601 gtaaatctggtacctgctcacaagaatttaacaagaagcagcatgattggcaaggct 660  
 Db 750 gtaaatctggtacctgctcacaagaatttaacaagaagcagcatgattggcaaggct 809

QY 661 ttactgaagagacagctcaagctattgagccatatattaccacttttttaacaggtt 720  
 Db 810 ttactgaagagacagctcaagctattgagccatatattaccacttttttaacaggtt 869

QY 721 ctgagcttgggaaacatctatcgagatttggtagagcatgtctttgacttaattttg 780  
 Db 870 ctgagcttgggaaacatctatcgagatttggtagagcatgtctttgacttaattttg 929

QY 781 gagctctacaattatgatagctatttgcctctctgtttttcccccagcttgaaatttaa 840

Db 930 gagctctacaattatgatagctatttgcctctctgttttccccagcttgaatttaa 989

QY 841 ttaaaagacaatgataaataagagagcgcctacaagttgttaaaactactgccaataatgttt 900

Db 990 ttaaaagacaatgataaataagagagcgcctacaagttgttaaaactactgccaataatgttt 1049

QY 901 ggggcaaaaggattcagaattggcttctcaaaaacaaagccactttggcagtgctacttgggc 960

Db 1050 ggggcaaaaggattcagaattggcttctcaaaaacaaagccactttggcagtgctacttgggc 1109

QY 961 aggtttaatgatccatgtaccatccgcctgggaatggtgaaaaatttctagcctattgt 1020

Db 1110 aggtttaatgatccatgtaccatccgcctgggaatggtgaaaaatttctagcctattgt 1169

QY 1021 ctcatgaaccatctctgatttagcaaaagacttaacagagatctttaaagtgggtcacat 1080

Db 1170 ctcatgaaccatctctgatttagcaaaagacttaacagagatctttaaagtgggtcacat 1229

QY 1081 gacctgggaagcctattagacatgatgttattgtgtcaatgttacagctgctcaaaaag 1140

Db 1230 gacctgggaagcctattagacatgatgttattgtgtcaatgttacagctgctcaaaaag 1289

QY 1141 gatattcttctgtccaatgacattacttaattttgtgagagagacattagacaaa 1200

Db 1290 gatattcttctgtccaatgacattacttaattttgtgagagagacattagacaaa 1349

QY 1201 cgtatggagtagtcgcaaaagacatgtagggacttggccaaatttataagaaaatgct 1260

Db 1350 cgtatggagtagtcgcaaaagacatgtagggacttggccaaatttataagaaaatgct 1409

QY 1261 ttacagtcagcagctggaaaagatgctgcacaaacagatagcatggatcaaaagacaaatcg 1320

Db 1410 ttacagtcagcagctggaaaagatgctgcacaaacagatagcatggatcaaaagacaaatcg 1469

QY 1321 ctacatatatttcaaaaatagattttagatgacgacttctgtgaaacggatctttgct 1380

Db 1470 ctacatatatttcaaaaatagattttagatgacgacttctgtgaaacggatctttgct 1529

QY 1381 caatacatggttctccacaaatttagaaactacagaaacgagatgaaatgcttattacttctg 1440

Db 1530 caatacatggttctccacaaatttagaaactacagaaacgagatgaaatgcttattacttctg 1589

QY 1441 tatgcacactggattttaaactgcttgaaagcattgaaatgctggaatgctggaatgctggaat 1500

Db 1590 tatgcacactggattttaaactgcttgaaagcattgaaatgctggaatgctggaatgctggaat 1649

QY 1501 ctgctccacacatcaagtaaaagatttgcctgacttgattaaagcaccacccaaacagatgcc 1560

Db 1650 ctgctccacacatcaagtaaaagatttgcctgacttgattaaagcaccacccaaacagatgcc 1709

QY 1561 agtgtcaaggcccatattttcaaaagtgatgggttattacaagaaatttacctgactcctgct 1620

Db 1710 agtgtcaaggcccatattttcaaaagtgatgggttattacaagaaatttacctgactcctgct 1769

QY 1621 aaggtcagagatttcatgaagaaatttcacacaggggtgttagaagagatgagagaaataaga 1680

Db 1770 aaggtcagagatttcatgaagaaatttcacacaggggtgttagaagagatgagagaaataaga 1829

QY 1681 aagcagttagaagtagtctgttagtccaacatgctcctcaagcaggtgaaaggttctgtg 1740

Db 1830 aagcagttagaagtagtctgttagtccaacatgctcctcaagcaggtgaaaggttctgtg 1889

QY 1741 cgtgaaataaactaaagaagtggggcaaccaccaacagcctacaaaatccttctctggaaatcg 1800

Db 1890 cgtgaaataaactaaagaagtggggcaaccaccaacagcctacaaaatccttctctggaaatcg 1949

QY 1801 atcaagtttctcttgagagagatagcacctgtgcacatagatcacggaatctctatcagtgct 1860

Db 1950 atcaagtttctcttgagagagatagcacctgtgcacatagatcacggaatctctatcagtgct 2009

QY 1861 ctatttaacaagtagtaacaaatcaatagatgaaacagcagatgagagatagaggttctt 1920



XX Homo sapiens.  
 OS Key Location/Qualifiers  
 PH exon 85..130  
 FT /tag= a  
 FT /number= 1  
 FT 131..257  
 FT /tag= b  
 FT /number= 2  
 FT 150..4325  
 FT /tag= c  
 FT /product= "AS3"  
 FT /note= "Androgen shutoff gene 3 protein"  
 FT 258..461  
 FT /tag= d  
 FT /number= 3  
 FT 462..548  
 FT /tag= e  
 FT /number= 4  
 FT 549..646  
 FT /tag= f  
 FT /number= 5  
 FT 647..773  
 FT /tag= e  
 FT /number= 6  
 FT 774..938  
 FT /tag= f  
 FT /number= 7  
 FT 939..995  
 FT /tag= g  
 FT /number= 8  
 FT 996..1111  
 FT /tag= h  
 FT /number= 9  
 FT 1112..1206  
 FT /tag= i  
 FT /number= 10  
 FT 1207..1352  
 FT /tag= j  
 FT /number= 11  
 FT 1353..1504  
 FT /tag= k  
 FT /number= 12  
 FT 1505..1618  
 FT /tag= l  
 FT /number= 13  
 FT 1619..1700  
 FT /tag= m  
 FT /number= 14  
 FT 1701..1749  
 FT /tag= n  
 FT /number= 15  
 FT 1750..1889  
 FT /tag= o  
 FT /number= 16  
 FT 1890..2005  
 FT /tag= p  
 FT /number= 17  
 FT 2006..2111  
 FT /tag= q  
 FT /number= 18  
 FT 2112..2272  
 FT /tag= r  
 FT /number= 19  
 FT 2273..2396  
 FT /tag= s  
 FT /number= 20  
 FT 2397..2555  
 FT /tag= t  
 FT /number= 21  
 FT 2556..2624  
 FT /tag= u

FT exon /number= 22  
 FT 2625..2761  
 FT /tag= v  
 FT /number= 23  
 FT 2672..2885  
 FT /tag= w  
 FT /number= 24  
 FT 2886..3090  
 FT /tag= x  
 FT /number= 25  
 FT 3091..3205  
 FT /tag= y  
 FT /number= 26  
 FT 3206..3338  
 FT /tag= z  
 FT /number= 27  
 FT 3339..3458  
 FT /tag= aa  
 FT /number= 28  
 FT 3459..3521  
 FT /tag= ab  
 FT /number= 29  
 FT 3522..3667  
 FT /tag= ac  
 FT /number= 30  
 FT 3668..3773  
 FT /tag= ad  
 FT /number= 31  
 FT 3774..4213  
 FT /tag= ae  
 FT /number= 32  
 FT 4214..4438  
 FT /tag= af  
 FT /number= 33  
 FT 4314..4408  
 FT /tag= ag  
 FT /note= "Destabilizing signal"  
 FT 4439..5337  
 FT /tag= ah  
 FT /number= 34  
 FT 5312..5317  
 FT /tag= ai  
 FT 5333..5337  
 FT /tag= aj  
 FT /note= "Poly adenylation cleavage site"  
 FT WO200050454-A1.  
 PN 31-AUG-2000.  
 PD 24-FEB-2000; 2000WO-US04732.  
 PF 24-FEB-1999; 99US-0121461.  
 PR (TUFT ) TUFTS COLLEGE.  
 PA Soto AM, Sonnenschein C, Geck P, Szelei J;  
 PI WPI; 2000-565451/52.  
 PP P-PSDB; AAY94702.  
 PS New human androgen-induced tumor suppressor cDNA sequence termed  
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient  
 PT diagnosis and treatment of prostate cancer -  
 XX Example 4; Fig 6; 152pp; English.  
 XX This invention relates to a human androgen-induced tumour suppressor  
 CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is  
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in  
 CC inhibiting cell proliferation and use as a marker for the efficient  
 CC diagnosis and treatment of prostate cancer. The invention includes AS3  
 CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host



```
QY 2221 gagaccagttgacagatatattgagcctctgcaataagagcctgagatcccaagcaacctg 2280
Db 2286 gagaccagttgacagatatattgagcctctgcaataagagcctgagatcccaagcaacctg 2345
QY 2281 gaacatcctaacaccattggtactattgctcatattgctctctctgacctgatcaa 2340
Db 2346 gaacatcctaacaccattggtactattgctcatattgctctctctgacctgatcaa 2405
QY 2341 ttgtgctccttggaatccttggttagctacttctcattgtgaaagatctcttcgatgaat 2400
Db 2406 ttgtgctccttggaatccttggttagctacttctcattgtgaaagatctcttcgatgaat 2465
QY 2401 gatcggttccagggaagaaagacaactaaactttgggttccagatgaagaaatctctct 2460
Db 2466 gatcggttccagggaagaaagacaactaaactttgggttccagatgaagaaatctctct 2525
QY 2461 gagacaaatggtcaaaatcaggtctataaattgattggttcgagtgctacttggaaatgaaa 2520
Db 2526 gagacaaatggtcaaaatcaggtctataaattgattggttcgagtgctacttggaaatgaaa 2585
QY 2521 aataatcacagtataatcagggaacttctaccttaagattgctaaacaaatatttgcatagt 2580
Db 2586 aataatcacagtataatcagggaacttctaccttaagattgctaaacaaatatttgcatagt 2645
QY 2581 gatgagacttgacagaaacagggaagaaattagtaaacacagatatgtcactgtgagactt 2640
Db 2646 gatgagacttgacagaaacagggaagaaattagtaaacacagatatgtcactgtgagactt 2705
QY 2641 gctgctggagtgctattgtgaagctggcacagaacccctgttacatgaaatcatcaca 2700
Db 2706 gctgctggagtgctattgtgaagctggcacagaacccctgttacatgaaatcatcaca 2765
QY 2701 ttagaacaaatcagctatgtcattagctatcaacgagatgaatgctatacaagtaagacaa 2760
Db 2766 ttagaacaaatcagctatgtcattagctatcaacgagatgaatgctatacaagtaagacaa 2825
QY 2761 ggtttgcccagaacttcaaaaagcccttcccggtttacggttccacctgagtatatg 2820
Db 2826 ggtttgcccagaacttcaaaaagcccttcccggtttacggttccacctgagtatatg 2885
QY 2821 gcaatctgtgcccctttgtgcaaaagatcctgttaaaggagagagagctcatctaggcaa 2880
Db 2886 gcaatctgtgcccctttgtgcaaaagatcctgttaaaggagagagagctcatctaggcaa 2945
QY 2881 tgtttggtgaaataataaattgaaggcgggtatcttgaaagcagcatgagctgttagt 2940
Db 2946 tgtttggtgaaataataaattgaaggcgggtatcttgaaagcagcatgagctgttagt 3005
QY 2941 gaaaaattattgctctctccagagtatgtgttccatatacaattcaaccttttggca 3000
Db 3006 gaaaaattattgctctctccagagtatgtgttccatatacaattcaaccttttggca 3065
QY 3001 catgaccagattatgtcaaatgacaggtatattgaacaaacttaagatgtttaaagaattg 3060
Db 3066 catgaccagattatgtcaaatgacaggtatattgaacaaacttaagatgtttaaagaattg 3125
QY 3061 ctttgggttcttggaataataatggtgctaaaaatgaaaaataacagtcacgtcttttacc 3120
Db 3126 ctttgggttcttggaataataatggtgctaaaaatgaaaaataacagtcacgtcttttacc 3185
QY 3121 aqaaagatggtagaataataaacaacaaagagatgcccaaggaccagatgatgcaaaa 3180
Db 3186 aqaaagatggtagaataataaacaacaaagagatgcccaaggaccagatgatgcaaaa 3245
QY 3181 atgaatgaaaaactgtacactgtgtgagttgccaatgaataatcatcatgttcaaaagagt 3240
Db 3246 atgaatgaaaaactgtacactgtgtgagttgccaatgaataatcatcatgttcaaaagagt 3305
QY 3241 actacatacagtttggaaatctctaaagaccgggtactaccagctcgtttcttccactcaa 3300
Db 3306 actacatacagtttggaaatctctaaagaccgggtactaccagctcgtttcttccactcaa 3365
QY 3301 cctgcaagaatttcagtaaacacccaaaaattatctgctcctctgaaatgaaatcatttttc 3360
```

```
Db 3366 cctgcaagaatttcagtaaacacccaaaaattatctgctcctgaaatgaaatcatttttc 3425
QY 3361 actcttgaaaaaccttaaaacacccaaatgttctagagagctgttaacaagccactttcatca 3420
Db 3426 actcttgaaaaaccttaaaacacccaaatgttctagagagctgttaacaagccactttcatca 3485
QY 3421 gcaggcaagcaatctcagaccacaaatcattacgaatggaactgtgaagcaatgcaagcagc 3480
Db 3486 gcaggcaagcaatctcagaccacaaatcattacgaatggaactgtgaagcaatgcaagcagc 3545
QY 3481 agctcaaatcgaagctcctctggaataaagggaggttgatagttctgaaatggat 3540
Db 3546 agctcaaatcgaagctcctctggaataaagggaggttgatagttctgaaatggat 3605
QY 3541 cacagtgaataatgaagattacacaatgtcttcacctttgcccggggaacaaagtgacaaag 3600
Db 3606 cacagtgaataatgaagattacacaatgtcttcacctttgcccggggaacaaagtgacaaag 3665
QY 3601 agagacgactctgattctgttaaggtctgaattggaagacccctagagggcaaaaaaacg 3660
Db 3666 agagacgactctgattctgttaaggtctgaattggaagacccctagagggcaaaaaaacg 3725
QY 3661 ccggtcacagaacagagaggaataataggtatggtgactgactaagttggtacagaa 3720
Db 3726 ccggtcacagaacagagaggaataataggtatggtgactgactaagttggtacagaa 3785
QY 3721 cagaaaccttaaaaggcagtcagcgaagtcggaaaaagggccatacagcttccagaatctgat 3780
Db 3786 cagaaaccttaaaaggcagtcagcgaagtcggaaaaagggccatacagcttccagaatctgat 3845
QY 3781 gaacgagctggtcctgaggaagaggtctcaaaagagatatattagaaaaatgaagatgaa 3840
Db 3846 gaacgagctggtcctgaggaagaggtctcaaaagagatatattagaaaaatgaagatgaa 3905
QY 3841 cagaatagtcgcgcaaaaaaggttaaaagagccgaccacccaaaaccttctgtggaggt 3900
Db 3906 cagaatagtcgcgcaaaaaaggttaaaagagccgaccacccaaaaccttctgtggaggt 3965
QY 3901 acaccaaaagagagcccaaatgaaacttctaaaaaagggaagcaaaaaaaacttgga 3960
Db 3966 acaccaaaagagagcccaaatgaaacttctaaaaaagggaagcaaaaaaaacttgga 4025
QY 3961 cctccagaccagagagaggaagaagaagaacaaaagtggaaatcgcgaacagaag 4020
Db 4026 cctccagaccagagagaggaagaagaagaacaaaagtggaaatcgcgaacagaag 4085
QY 4021 tccaaaagcaaacacagcagcagtgctcaaggagagcacagcagagcagaatctctctgaa 4080
Db 4086 tccaaaagcaaacacagcagcagtgctcaaggagagcacagcagagcagaatctctctgaa 4145
QY 4081 tctagtcaattgaatccacacagctccacacacagaaaggaggaagacccatcaaaa 4140
Db 4146 tctagtcaattgaatccacacagctccacacacagaaaggaggaagacccatcaaaa 4205
QY 4141 acgcatcacctatcacaccccaaaaaaaatgtg 4173
Db 4206 acgcatcacctatcacaccccaaaaaaaatgtg 4238
```

## RESULT 2

AAA28052

ID AAA28052 standard; cDNA; 5355 BP.

XX AAA28052;

XX AAA28052;

DT 01-DEC-2000 (first entry)

XX Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.

DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;

XX chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;

KW diagnosis; treatment; cytostatic; human; ss.



Db 66 atggctcattcaagactaggaccatgatggaaaaattacatatccgctgggccaag 125  
QY 61 gaatatcagataaataataaagaggagatggtcagacgataaagatggttgaaa 120  
Db 126 gaaatacagataaataataaagaggagatggtcagacgataaagatggttgaaa 185  
QY 121 actttatgatgatgaccagagactctgaagaagaaagagctttattttaaacctagct 180  
Db 186 actttatgatgatgaccagagactctgaagaagaaagagctttattttaaacctagct 245  
QY 181 ttacatttcttcagatttttttctcagacatccctggtaaaagatttcttactggta 240  
Db 246 ttacatttcttcagatttttttctcagacatccctggtaaaagatttcttactggta 305  
QY 241 gctgctgcttctgatatatttcagagatttatgctcctgaagctcttacacatccct 300  
Db 306 gctgctgcttctgatatatttcagagatttatgctcctgaagctcttacacatccct 365  
QY 301 gataaactaaagatatattttatgtttaaacagacagttgaagggtctagaggataca 360  
Db 366 gataaactaaagatatattttatgtttaaacagacagttgaagggtctagaggataca 425  
QY 361 aagagcccaaatcctaagatgttatttttatttacttgagaacattgctgggtcaagtc 420  
Db 426 aagagcccaaatcctaagatgttatttttatttacttgagaacattgctgggtcaagtc 485  
QY 421 tataacatatgctttgatttagaagatagcaatgaaattttaccagctatacagaacc 480  
Db 486 tataacatatgctttgatttagaagatagcaatgaaattttaccagctatacagaacc 545  
QY 481 ttattttcagttataaacaatggccacaaacagaaagtcctatgcacatgggtagacctt 540  
Db 546 ttattttcagttataaacaatggccacaaacagaaagtcctatgcacatgggtagacctt 605  
QY 541 atgagctctattttgtgaagtgatatacagtgctctcgaagagcttttgatacggtttta 600  
Db 606 atgagctctattttgtgaagtgatatacagtgctctcgaagagcttttgatacggtttta 665  
QY 601 gtaaatctgtctactctcaataagaatttaaacaaagcaacatataatgttgcaaggct 660  
Db 666 gtaaatctgtctactctcaataagaatttaaacaaagcaacatataatgttgcaaggct 725  
QY 661 ttactgaagaggacagctcaagctatttgagccataattaccacttttttaacaggtt 720  
Db 726 ttactgaagaggacagctcaagctatttgagccataattaccacttttttaacaggtt 785  
QY 721 ctgatgcttgggaaacacatacagcgtattgtcagagacatgtcttgacttaatttg 780  
Db 786 ctgatgcttgggaaacacatacagcgtattgtcagagacatgtcttgacttaatttg 845  
QY 781 gagctctacaataattgatagctatttgctgctctctgttttaccagcttgaaatttaa 840  
Db 846 gagctctacaataattgatagctatttgctgctctctgttttaccagcttgaaatttaa 905  
QY 841 ttaaaagagcaatgataatgaggagcgcctacaaagtgtttaaaactactggcaaaaatgttt 900  
Db 906 ttaaaagagcaatgataatgaggagcgcctacaaagtgtttaaaactactggcaaaaatgttt 965  
QY 901 ggggcaagagattcagaatttggcttctcaaaaacagccactttggcagtgctacttgggc 960  
Db 966 ggggcaagagattcagaatttggcttctcaaaaacagccactttggcagtgctacttgggc 1025  
QY 961 aggtttaatgatcatcatgtaccatccgctggaaattgtgaaatttctagccattgt 1020  
Db 1026 aggtttaatgatcatcatgtaccatccgctggaaattgtgaaatttctagccattgt 1085  
QY 1021 ctcatgaaccatcctgatttagcaaaagacttaacagagatctttaaaagtgaagtcacat 1080  
Db 1086 ctcatgaaccatcctgatttagcaaaagacttaacagagatctttaaaagtgaagtcacat 1145  
QY 1081 gacctgaggaagctattagacatgatgttattgtgtcaatagttacagctgtctaaaaag 1140  
Db 1146 gacctgaggaagctattagacatgatgttattgtgtcaatagttacagctgtctaaaaag 1205

QY 1141 gatattcttctggtcgaatgatcacttacttaattttgtgagagagagaacattagacaaa 1200  
Db 1206 gatattcttctggtcgaatgatcacttacttaattttgtgagagagagaacattagacaaa 1265  
QY 1201 cgatggagagtacgcaagaagccatgatgggacttggccaaattttataagaataatgct 1260  
Db 1266 cgatggagagtacgcaagaagccatgatgggacttggccaaattttataagaataatgct 1325  
QY 1261 ttacagtcagcagctggaagaagatgctgcaaaaacagatagctggatcaaaagcaaatg 1320  
Db 1326 ttacagtcagcagctggaagaagatgctgcaaaaacagatagctggatcaaaagcaaatg 1385  
QY 1321 ctacatatatttatcaaaatagatttattgatgcgactacttgttgaacgagctttgct 1380  
Db 1386 ctacatatatttatcaaaatagatttattgatgcgactacttgttgaacgagctttgct 1445  
QY 1381 caatacatggttctcctcacaaatttagaaactacagaacggatgaaattgcttatattactg 1440  
Db 1446 caatacatggttctcctcacaaatttagaaactacagaacggatgaaattgcttatattactg 1505  
QY 1441 tatgcacactggattttaaagtctgtgaagcattgaatgaaatgtggaatgtcaaaaat 1500  
Db 1506 tatgcacactggattttaaagtctgtgaagcattgaatgaaatgtggaatgtcaaaaat 1565  
QY 1501 ctgctccgacatcaagtaaaagatttctgacttgaatgaaccccaaaacagatgcc 1560  
Db 1566 ctgctccgacatcaagtaaaagatttctgacttgaatgaaccccaaaacagatgcc 1625  
QY 1561 agtgtcaaggccatattttcaaaagtgtgttatcaagaataattacctgactcctggt 1620  
Db 1626 agtgtcaaggccatattttcaaaagtgtgttatcaagaataattacctgactcctggt 1685  
QY 1621 aaggctcagagatttctgaagaataatcacaaggtttagaagatgatgagaataaaga 1680  
Db 1686 aaggctcagagatttctgaagaataatcacaaggtttagaagatgatgagaataaaga 1745  
QY 1681 aagcagttagaagtacttctcacaacatgctcctgcagcaggtgaaggtgtgtg 1740  
Db 1746 aagcagttagaagtacttctcacaacatgctcctgcagcaggtgaaggtgtgtg 1805  
QY 1741 cgtgaaataactaagaagtgtgggcaaccccaaaacagcctacaatcctttcctggaaatg 1800  
Db 1806 cgtgaaataactaagaagtgtgggcaaccccaaaacagcctacaatcctttcctggaaatg 1865  
QY 1801 atcaagtttcttggagagatagcacctgtgcacatagataccgaaatctatcagtgct 1860  
Db 1866 atcaagtttcttggagagatagcacctgtgcacatagataccgaaatctatcagtgct 1925  
QY 1861 ctatttaacaagaatgaacaaatcaatagatggaaacagcagatgatgaagatgaggtgtt 1920  
Db 1926 ctatttaacaagaatgaacaaatcaatagatggaaacagcagatgatgaagatgaggtgtt 1985  
QY 1921 ccaactgataagccatcagagcaggtcttgaactgcttaaggtactctcattacacat 1980  
Db 1986 ccaactgataagccatcagagcaggtcttgaactgcttaaggtactctcattacacat 2045  
QY 1981 cccatctcatttctctgctgaaacatttgaatcattactgctgtctgaaatggat 2040  
Db 2046 cccatctcatttctctgctgaaacatttgaatcattactgctgtctgaaatggat 2105  
QY 2041 gatgaaaaagtagcagaagctgcactacaattttcaaaaacacaggaagcaaaattgaa 2100  
Db 2106 gatgaaaaagtagcagaagctgcactacaattttcaaaaacacaggaagcaaaattgaa 2165  
QY 2101 gaggattttccacacatcagatcagccttctcctgtttttacatcacaaatctaaaaa 2160  
Db 2166 gaggattttccacacatcagatcagccttctcctgtttttacatcacaaatctaaaaa 2225  
QY 2161 gggccccccgtcaagccaaatagtcattcattgtatccatcgatattttctagtaa 2220  
Db 2226 gggccccccgtcaagccaaatagtcattcattgtatccatcgatattttctagtaa 2285



```
FT exon 465..562
FT /*tag= f
FT /number= 5
FT exon 563..689
FT /*tag= e
FT /number= 6
FT exon 690..854
FT /*tag= f
FT /number= 7
FT exon 855..911
FT /*tag= q
FT /number= 8
FT exon 912..1027
FT /*tag= h
FT /number= 9
FT exon 1028..1122
FT /*tag= i
FT /number= 10
FT exon 1123..1268
FT /*tag= j
FT /number= 11
FT exon 1269..1420
FT /*tag= k
FT /number= 12
FT exon 1421..1534
FT /*tag= l
FT /number= 13
FT exon 1535..1616
FT /*tag= m
FT /number= 14
FT exon 1617..1665
FT /*tag= n
FT /number= 15
FT exon 1666..1805
FT /*tag= o
FT /number= 16
FT exon 1806..1921
FT /*tag= p
FT /number= 17
FT exon 1922..2027
FT /*tag= q
FT /number= 18
FT exon 2028..2188
FT /*tag= r
FT /number= 19
FT exon 2189..2312
FT /*tag= s
FT /number= 20
FT exon 2313..2471
FT /*tag= t
FT /number= 21
FT exon 2472..2540
FT /*tag= u
FT /number= 22
FT exon 2541..2677
FT /*tag= v
FT /number= 23
FT exon 2678..2801
FT /*tag= w
FT /number= 24
FT exon 2802..3006
FT /*tag= x
FT /number= 25
FT exon 3007..3121
FT /*tag= y
FT /number= 26
FT exon 3122..3254
FT /*tag= z
FT /number= 27
FT exon 3255..3374
FT /*tag= aa
FT /number= 28
FT 3375..3437
```

```
FT exon /*tag= ab
FT /number= 29
FT 3438..3583
FT /*tag= ac
FT /number= 30
FT exon 3584..3689
FT /*tag= ad
FT /number= 31
FT exon 3690..4129
FT /*tag= ae
FT /number= 32
FT exon 4130..4354
FT /*tag= af
FT /number= 33
FT misc_signal
FT /*tag= ag
FT /note= "Destabilizing signal"
FT exon 4355..5253
FT /*tag= ah
FT /number= 34
FT polyA_signal
FT 5228..5233
FT /*tag= ai
FT misc_feature
FT 5249..5253
FT /*tag= aj
FT /note= "Poly adenylation cleavage site".
FT
XX
PN WO200050454-A1.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-US04732.
XX
XX 24-FEB-1999; 99US-0121461.
XX (TUFT ) TUFTS COLLEGE.
XX
XX Soto AM, Sonnenschein C, Geck P, Szelei J;
XX
XX WPI; 2000-565451/52.
XX P-PSDB; AAY94702.
XX
XX New human androgen-induced tumor suppressor cDNA sequence termed
XX 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX diagnosis and treatment of prostate cancer -
XX
XX Claim 1; Fig 1; 152pp; English.
XX
XX This invention relates to a human androgen-induced tumour suppressor cDNA
XX sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
XX on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
XX proliferation and use as a marker for the efficient diagnosis and
XX treatment of prostate cancer. The invention includes AS3 cDNA and protein
XX sequences, a vector comprising the cDNA sequence, a host cell transfected
XX with the expression vector, and a method for producing an AS3 polypeptide
XX comprising culturing the transfected cells. AS3 has cytostatic activity,
XX and acts to suppress cell proliferation. The AS3 gene is useful as a
XX marker for the efficient diagnosis and treatment of prostate cancer. The
XX AS3 nucleic acid molecule can be used as a source of antisense agents for
XX sequence specific modulation of gene expression. The AS3 protein may be
XX used in the treatment of disorders caused by aberrant modification or
XX mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
XX or aberrant post-translational modification of the AS3 protein. This
XX sequence represents the human AS3 cDNA sequence.
XX
XX Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;
```

Query Match 100.0%; Score 4173; DB 21; Length 5271;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctattcaagactagaccatgatggaaaattacatatccgctgggtcaag 60  
|||||



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 19:52:23 ; Search time 520.53 seconds  
(without alignments)  
5033.780 Million cell updates/sec

Title: US-09-512-581-3

Perfect score: 4173

Sequence: 1 atggctcattcaagactag.....cacaacccaataaaatgtg 4173

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

N\_Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4173	100.0	5271	21	AAA28051 Human androgen shu
2	4173	100.0	5355	21	AAA28052 Human androgen shu
3	1112.8	26.7	3957	21	AAA47423 Sequence encoding
4	411.4	9.9	439	20	AAV87629 EST clone DY17. H
5	348	8.3	738	20	AAZ15259 Human gene express
6	298.8	7.2	351	21	AAZ42861 Human 5' EST isola
7	292	7.0	295	21	AAC02817 Human secreted pro
8	253.8	6.1	1217	21	AAF18338 Lung cancer associ
9	203.6	4.9	530	21	AAZ80598 Human colon cancer
10	133.2	3.2	936	22	AAF58252 Oligonucleotide D1
11	133.2	3.2	936	22	AAF58254 Oligonucleotide D1

c 12	133.2	3.2	936	22	AAF58257	Oligonucleotide D1
c 13	133.2	3.2	936	22	AAF58259	Oligonucleotide D2
c 14	133.2	3.2	936	22	AAF58262	Oligonucleotide D2
c 15	133.2	3.2	938	22	AAF58255	Oligonucleotide D1
c 16	132.4	3.2	936	22	AAF58252	Oligonucleotide D1
17	132.4	3.2	936	22	AAF58254	Oligonucleotide D1
18	132.4	3.2	936	22	AAF58257	Oligonucleotide D1
19	132.4	3.2	936	22	AAF58259	Oligonucleotide D2
20	132.4	3.2	936	22	AAF58262	Oligonucleotide D2
21	132.4	3.2	938	22	AAF58255	Oligonucleotide D1
22	99	2.4	161	21	AAC06611	Human secreted pro
23	59.2	1.4	6644	20	AAZ33181	Base sequence of t
24	59.2	1.4	7372	20	AAZ33182	Base sequence of t
25	59.2	1.4	7797	20	AAZ33180	Cowpox virus bsr f
26	59.2	1.4	7996	20	AAZ33184	Base sequence of t
27	57.6	1.4	4000	18	AAZ191902	Mannose-1-phosphat
c 28	54.8	1.3	244	22	AAF58238	Oligonucleotide D1
29	54.6	1.3	6755	19	AAV21511	staphylococcal bac
30	54.4	1.3	244	22	AAF58238	Oligonucleotide D1
c 31	53	1.3	5059	20	AAZ84332	stealth virus nucl
c 32	51.6	1.2	3399	17	AAZ05868	Chicken leucocytoz
c 33	51.4	1.2	72604	20	AAZ10752	Genomic sequence o
c 34	48.6	1.2	396	22	AAZ94862	Human ovarian can
c 35	48.2	1.2	1236	21	AAZ02163	Human colon cancer
36	48.2	1.2	10732	21	AAZ10594	Gene encoding a su
c 37	48.2	1.2	49959	20	AAZ23895	Murine LOBO homolo
38	47.6	1.1	6741	21	AAZ10595	Gene encoding a su
39	47.2	1.1	1686	16	AAZ07587	DNA encoding leuco
40	45.8	1.1	49999	20	AAZ23891	Murine LOBO homom
41	45.8	1.1	49999	20	AAZ23896	Murine LOBO homom
42	45.6	1.1	979	20	AAZ61816	B. burgdorferi ant
43	45.6	1.1	1083	20	AAZ61815	B. burgdorferi ant
44	45.6	1.1	2487	20	AAZ20298	Borrelia burgdorfe
45	45.2	1.1	43576	21	AAZ68609	Bacteriophage 96 c

## ALIGNMENTS

RESULT 1  
AAA28051  
ID AAA28051 standard; cDNA; 5271 BP.  
XX  
AC AAA28051;  
XX  
DT 01-DEC-2000 (first entry)  
XX  
DE Human androgen shutoff gene 3 (AS3) cDNA sequence.  
XX  
KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;  
KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;  
KW diagnosis; treatment; cytostatic; human; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT exon  
FT Location/Qualifiers  
FT 1..46  
FT /\*tag= a  
FT /number= 1  
FT 47..173  
FT /\*tag= b  
FT /number= 2  
FT 66..4241  
FT /\*tag= c  
FT /product= "AS3"  
FT /note= "Androgen shutoff gene 3 protein, the CDS is specifically claimed as SEQ ID #3"  
FT 174..377  
FT /\*tag= d  
FT /number= 3  
FT 378..464  
FT /\*tag= e  
FT /number= 4



Wed Sep 26 11:24:08 2001

us-09-512-581-3.rst

Page 14



REFERENCE 1 (bases 1 to 888)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: L10CM1558 row: P column: 11  
High quality sequence stop: 688.

## FEATURES

Source

1. 888  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4714234"  
/clone\_id="NIH\_MGC\_79"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);  
Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggcattatggc)  
); 5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCCGAGCGCCGACATG-dr(30)BN-3'  
(where B = A, C, G, or T). Average  
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."

BASE COUNT 347 a 163 c 223 g 149 t  
ORIGIN

Query Match 11.2%; Score 468; DB 155; Length 888;  
Best Local Similarity 99.8%; Pred. No. 3.4e-103;  
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3693 ggaatgaactgaactaagtgtgtacaggaacagaacaaacctaaggcagtcagcgaagtcgga 3752  
Dd 1 GGATGACTTGACTGACTAGTGTGACAGAACAGAACCTTAAGGCACTCAGCGAAGTCGGA 60  
QY 3753 aagagggcctacggcttcagaatctgtatgaacagcagctggcctgaggaagaagagctcaa 3812  
Dd 61 AAGAGGCCATCGCGCTTCAGATCTGATGAACAGCAGTGGCTTGAGAAAGAGCGCTCAA 120  
QY 3813 agaagatatattagaanaatgaagatgaacagaatagtcgcgcaaaaaaagggttaaaagg 3872  
Dd 121 AGAAGATATATTAGAAATGTAAGATGAACAGATATGTCGCC-AAAAAGGTTAAAGAGG 179  
QY 3873 ccgaccacaaaacctctgtgtgagtgacacaaagaagaagccaacatgaaaacttc 3932  
Dd 180 CCGACACACCAAAACCTTGTGTGAGGTACACCAAAAGAGAGCCACACATGAAAACTTC 239  
QY 3933 taataaagaagaacaaaaaatacttggaactccagcaccagagaagaagaagaaga 3992  
Dd 240 TAAAAAAGAGCAAAAAAATCTGACCTCCAGCACAGAGAGGAAGGAAGA 299  
QY 3993 aagacaagtggaatacgaacagaagaagtcacaaagcaaacagcagagtgtaaggag 4052  
Dd 300 AAGACAAAGTGGAATACGGAACAGAAAGTCCAAAAGCAACAGCAGGTCTCAAGGAG 359  
QY 4053 agcacaagcagagcagaatctcctgaatctagtgaattgaatccacacagtcacacc 4112  
Dd 360 AGCACAGCAGAGAGAGATCTCTGAATCTAGTCAATTGAAATCCACACAGTCCACACC 419  
QY 4113 acagaagaagcaggaagaacacatacaaaaagcaccatcacatcaacaacaaaaaatgt 4172  
Dd 420 ACAGAAAGAGAGAGAGAACCATCAAAAAAGCCATCACCATCACAACCAAAAAAATGT 479

Search completed: September 25, 2001, 16:19:35  
Job time: 6023 sec



61 gaataatcagataaataatcctaaagagagatggtgagacgattaaagatggtgtaa 120  
 110 GAAATCTCAGATAAATCTCTAAAGAGAGATGCTGAGCGGCTTAAAGATGTTGTA 169  
 121 accttaatgatatgacccagagacttgaagaagaagaagacttattaaaccagct 180  
 170 ACTTATGAGACATGAGACAGACCTGTGAAGAGAAAGAACTTATCTTAAACCTAG 229  
 181 ttaacatctgctcagatcttctcctcaagcactctgtaagaatgcttgcctacgt 240  
 230 TTACATCTTCTCTCTACTCTTCTCTCAAGCATCTGATTAAGATGCTTACTG 289  
 241 gctctgctgctgctgataatcctcaagattatgctcctgaagctccttaacaccc 300  
 290 GCTTGTGCTGCTGCTGATATTTTTCAGGATTTATGCTCTGAGCTCTCTTACG 349  
 301 gataaactaagaagatatattatgcttataaagaacagtgtaaggagctagaagata 360  
 350 GATTAACCTAAGAGATATTTATGTTATTAACAGGCACTTAAGGAGACTAGAA 409  
 361 aagagcccaatcattcagatgatttatttacttgaagacattgctggtcaagta 420  
 410 AAGAGCCCTCAATTAATTAATGTTATTTTATTTACTGAGAACATTCGATGGTT 469  
 421 tataacatactgcttgaatgaagaatgaagaatgaagaatgaagaatgaagaatga 480  
 470 TATAACATATGCTTGTGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 529  
 481 ttatttcaatgaataaagaatgagacacacacacacacacacacacacacacac 540  
 530 TTAATTCGATTAATTAACATGAGACATGAGATGAGATGAGATGAGATGAGATG 589  
 541 atgagctcattatctgtg 559  
 590 ATGAGCTCTATCATTTGTG 608

**RESULT 14**  
 LOCUS B511856 490 bp mRNA EST 06-DEC-2000  
 DEFINITION UI-H-B14-abs-d-12-0-UI-s1 NCI-CGAP-Sub8 Homo sapiens CDNA clone  
 IMAGE:308858 3', mRNA sequence.  
 ACCESSION B511856  
 VERSION B511856.1 GI:11595154  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 490)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 The sequence contained an oligo-dt track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. CDNA Library Preparation: M.B. Soares Lab Clone distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbfp/image/image.html  
 Seq primer: M13 Forward  
 POLYA=yes.

**FEATURES**  
 Source  
 1. 490  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:308858"  
 /clone\_lib="NCI CGAP Sub8"  
 /lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; NCI-CGAP-Sub8  
 is a subtracted library derived from NCI-CGAP-Sub5. The  
 NCI-CGAP-Sub8 library had 2.5 million recombinants. A  
 single-stranded DNA preparation of NCI-CGAP-Sub5 was used  
 as a tracer in a subtractive hybridization with a driver  
 comprising: a pool of clones from NCI-CGAP-Sub5 (IMAGE  
 clone ids 2732833-2737415, 3068040-3069191; 25% of the  
 driver population), a pool of clones from NCI-CGAP-Sub4  
 (IMAGE clone ids 2723592-2729326; 25% of the driver  
 population), NCI-CGAP-Sub6 (pool AIF-AU, IMAGE ids  
 2728969-2733190; 25% of the driver population), and  
 NCI-CGAP-Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550  
 ; 25% of the driver population). Subtraction was  
 performed as previously described (Bonafide, Lennon &  
 Soares (1996): Normalization and subtraction. Two  
 Approaches To Facilitate Gene Discovery. Genome Research  
 6, 791-806.  
 TAG\_LIB=NCI-CGAP\_Lu5  
 TAG\_TISSUE=Lung  
 TAG\_SEO=CAAC"

BASE COUNT 167 a 79 c 96 g 148 t  
 ORIGIN  
 Query Match 11.4%; Score 474.8; DB 149; Length 490;  
 Best local Similarity 99.6%; Pred. No. 6.4e-105;  
 Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2680 tgltaacatgaatcattacattagaacaatcagctatgcatgaatcagat 2739  
 DB 13 TGTATACATGAATATCATCATATGAGAACATATACATGATGCTATGATCAACGAT 72  
 QY 2740 gaatgctatcaatgaagaagaagtggtgcccagaactcagaagaagccttcgatta 2799  
 DB 73 GAATGCTATCAATGAAGACAGAGATGTTGCCAGAACTTCACAAAGCCCTTCCGTTA 132  
 QY 2800 cggcttcacattgagatataatgcaatctgccccttctgcaagaagatcctgtaagag 2859  
 DB 133 CGGCTTCCACTGATGATATGATGATGATGATGATGATGATGATGATGATGATG 192  
 QY 2860 agagaagctcattctgctgcaatgcttctgcaagaatataatgtaagggagatctg 2919  
 DB 193 AGAAGAGCTCATCTGAGCAATGTTGGTCMAAAATATAATATGAGGCGGATCTG 252  
 QY 2920 aagcagcatgcaagcttctgcaagaatataatgctccttctcaccagagatgcttcca 2979  
 DB 253 AACGACATGACCTGTTAGTGAATAATTTTGTCTCTTACACAGATGTTGTTCCA 312  
 QY 2980 tatacaattcacccttctgcaatgacccagatattatgcaagtagaagattgacaa 3039  
 DB 313 TATACATATTCACCTTTGGACATGACCCAGATGATGTAAGATGATGTAACAA 372  
 QY 3040 cctaaagatgtaagaagaatgcttctgctgctgcaagaatataatgtaagggag 3099  
 DB 373 CTAAAGAGATGAAGAAAGTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 432  
 QY 3100 aataaagctcagcttctcagagaagatgtaagaatataatgtaagaagaatg 3157  
 DB 433 AATAACAGTCACCTGTTATGACAAAGATGTTGAAATAATTTAAACAAACAAAGATG 490

**RESULT 15**  
 LOCUS BG571128 888 bp mRNA EST 10-APR-2001  
 DEFINITION 602591748F1 NIH\_MCC\_79 Homo sapiens CDNA clone IMAGE:4714234 5',  
 mRNA sequence.  
 ACCESSION BG571128  
 VERSION BG571128.1 GI:13578781  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution by: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LINL at:  
 www-bio.linnl.gov/bdrp/image/image.html  
 Insert length: 1048 Std Error: 0.00  
 Seq primer: -40UP from Gldco  
 High quality sequence stop: 455.

## FEATURES

Source  
 1..480  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2239331"  
 /clone\_lib="NCI CGAP GC6"  
 /tissue-type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 82 a 123 c 84 g 191 t  
 ORIGIN

Query Match 11.5%; Score 478.4; DB 23; Length 480;  
 Best Local Similarity 99.8%; Pred. No. 8.4e-106;  
 Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3460 acgtgaagaatgcgaagcagctcaatccaaagctctcttggaagaataaaggagg 3519  
 |||||||  
 480 ACTGTAACCAATGCAGAGCAGCTCAATCCAGCTCTCTGGAAGATAAAGGGAGG 421  
 |||||||  
 3520 cttaagatctgaagaatgcagtgaaatgaagaattcacatctctacccttg 3579  
 |||||||  
 420 CTGATAGCTTCTGAAATGATCAGTGAATAAGATACACATGCTTCCACCTTGG 361  
 |||||||  
 3580 ccggggaaaaaagtgcagaagagacactctgactgtaagctctgaattgagaag 3639  
 |||||||  
 360 CCGGGGAAAAAAGTGCAGAGACACACTCTGATCTTGTAGGCTCTGAATTGAGAAG 301  
 |||||||  
 3640 cctgaagcagaagaaaaaacgcccgtcaacagacagagagagaaattagatgac 3699  
 |||||||  
 300 CCTGAGAGCAGGAAAAAAGCGCCCTCACAGAACAGAGAGAGAAATTAGATGATGAC 241  
 |||||||  
 3700 ttgactaagtggtgataagaagaagaacctaagcagctcagcgaagtcggaagaagc 3759  
 |||||||  
 240 TTGACTAAGTGTGATACAGACAGAAACCTTAAAGGACATCAGCAAGTCGGAAGAGCG 181  
 |||||||  
 3760 catacagctcagaatctgcagacagcagtgctgaggaagaagagctcaagaagat 3819  
 |||||||  
 180 CATACGCTTCAAGATCTGATGATGACAGCATGTGCTGAGGAAAAAGAGCTCAAAAGAT 121  
 |||||||  
 3820 atattagaanaatgaagatgcagaagaatagtcgcgcaaaaaaggttaaaagagccgacca 3879  
 |||||||  
 120 ATATTAGAANAATGAAGATGACAGCATAGTCCGCCAAAAAAGGTTAAAGAGAGCGCAGCA 61  
 |||||||  
 3880 ccaaaacctctgtgtggtggtgatacacaagaagaagccacaatgaataactctaaaaa 3939  
 |||||||  
 60 CCAAAACCTCTGTGTGGTGTACACCAAAAGAGCCAAAAATGATAAACTCTTAAAAA 1  
 |||||||

RESULT 13  
 LOCUS BF471938 608 bp mRNA EST 04-DEC-2000  
 DEFINITION UI-M-BH3-avr-g-03-0-UI.r1 NIH\_BMAP\_M.S4 Mus musculus cDNA clone  
 UI-M-BH3-avr-g-03-0-UI 5', mRNA sequence.

ACCESSION BF471938  
 VERSION BF471938.1 GI:11541121  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 608)  
 REFERENCE  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704447  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@mail.nih.gov

## FEATURES

Source

Location/Qualifiers  
 1..608  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-avr-g-03-0-UI"  
 /clone\_lib="NIH\_BMAP\_M.S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH\_BMAP\_M.S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M.S4, NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2, NIH\_BMAP\_M.S3.1, NIH\_BMAP\_M.S2, NIH\_BMAP\_M.S1. The subtracted library (NIH\_BMAP\_M.S4) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2, and NIH\_BMAP\_M.S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2, and NIH\_BMAP\_M.S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, into DH10B bacteria (Life Technologies) and electropotated into NIH\_BMAP\_M.S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)."  
 BASE COUNT 193 a 113 c 128 g 174 t  
 ORIGIN

Query Match 11.4%; Score 477.4; DB 149; Length 608;  
 Best Local Similarity 90.9%; Pred. No. 1.6e-105;  
 Matches 508; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

1 atggtcatctcaagaactagacaaatgataaatacatcgcctgggtgaag 60  
 |||||||  
 50 ATGGCTCATTTCAAGACCAAGCATGGGAAATTAATTACTTACCCTCGAATCAAG 109  
 |||||||







AV685306  
 LOCUS AV685306 608 bp mRNA EST 25-SEP-2000  
 DEFINITION AV685306 GKC Homo sapiens cDNA clone GKATF08 5', mRNA sequence.  
 ACCESSION AV685306  
 VERSION AV685306.1 GI:10287169  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 608)  
 AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.  
 TITLE Homo sapiens cDNA GK- clones  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn  
 This clone is available at CHGC in Shanghai.  
 FEATURES  
 source 1. 608  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GKATF08"  
 /clone\_id="GKC"  
 /tissue\_type="hepatocellular carcinoma"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 181 a 113 c 125 g 189 t  
 ORIGIN  
 Query Match 12.1%; Score 503.8; DB 32; Length 608;  
 Best Local Similarity 96.2%; Pred. No. 5.7e-112;  
 Matches 528; Conservative 0; Mismatches 17; Indels 4; Gaps 1;  
 385 tttaattactgaagaacattgcttggtgacatgacataataatgcttgaattgaa 444  
 |||||  
 2 ttttatttacttgagAACATTGCTTGGGCAAGGCATATGCTTGTAGTTAGAA 61  
 445 gatagaacatgaattttaccacagctatacagaacctattttcagttataacaatg 504  
 |||||  
 62 GATAGCAATGAATTTTACCAGCTATACAGAACCTTATTTCAGTTATAACAATGGC 121  
 505 cacaatcagaagtcacatgacatgtagaagccttaagagctcattattgtgaagt 564  
 |||||  
 122 CACATTCAGAAAGTCCATATGACATGCTAGACCTTATGAGCTCTATTATTGTGAAG 181  
 565 gatacagctcagaagcctttgtgatacagtttaagtaaatctgtaacctcacaag 624  
 |||||  
 182 GATACAGTGTCTCAAGAGCTTTTGATACGATTTAGTAATCTGTGACCTCATAG 241  
 625 aatttaacaagaacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 684  
 |||||  
 242 AATTTAACCAAGCAAGCATATGATTGGCAAGGCTTACTGAAGAGGACACCTCAAG 301  
 685 attgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 744  
 |||||  
 302 ATTGAGCATATATACCAATTTTATATCAGATCTGATGCTTGGGAAACATCTATC 361  
 745 agcgattgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 804  
 |||||  
 362 AGCGATTTGTGAGCAGCATGCTTGTGACTTAATTGAGAGCTCTTACAAATATGATG 421  
 |||||  
 805 ttgcgtcctctgttttaaccagcgttgatttaataagaagcagcagcagcagcagc 864

|||||  
 Db 422 TTGCTGCTCTCTGTTTACCAGCTTGAATTAACATAAGAGCAATGATATGAGAG 481  
 Oy 865 cgcctacaagtggttaactactgacaaaatggttggcgcaaaagatcgaattgct 924  
 |||||  
 Db 482 CGCCTTCAAGCTGATTAACCTACTGGC---ATGTTAGTGGCAAGGATTCAGAACTGGCT 537  
 Oy 925 tctcaaac 933  
 |||||  
 Db 538 TCTCCATC 546  
 RESULT 10  
 LOCUS BE532986 683 bp mRNA EST 09-AUG-2000  
 DEFINITION 601235013P1 NCL\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3598922 5', mRNA sequence.  
 ACCESSION BE532986  
 VERSION BE532986  
 KEYWORDS BE532986.1 GI:9761631  
 SOURCE EST.  
 ORGANISM house mouse.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 683)  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-rt@mail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNI)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNI at:  
 http://image.llnl.gov  
 Plate: LLM8780 row: 1 column: 03  
 High quality sequence stop: 607.  
 FEATURES  
 source 1. 683  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3598922"  
 /clone\_id="NCL\_CGAP\_Mam6"  
 /sex="Female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: PCMV-SPORE6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies, Inc. Investigator providing samples: Jeffrey Green, M.D., NIH"  
 BASE COUNT 235 a 146 c 138 g 163 t  
 ORIGIN  
 Query Match 11.9%; Score 495; DB 136; Length 683;  
 Best Local Similarity 88.6%; Pred. No. 8.3e-110;  
 Matches 559; Conservative 0; Mismatches 70; Indels 2; Gaps 2;  
 Oy 2930 cagcgttagtgaataatat-tgctctctcacaagtgatgtgtccatatacaatt 2988  
 |||||  
 Db 1 CAGCTGTAGTGAATAAATATATGTCTCTTACCAAGATATGTTCCATATACAAAT 60  
 Oy 2989 cacccttggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3048  
 |||||  
 Db 61 CACCTTTGGCAGACAGCCACATATATGCAAGATACAGAGATATGACCACTTAAAGT 120  
 Oy 3049 gttaaagaatgctcttggttggttcttggaataatgaatgcttaaaatgaataacagt 3108  
 |||||  
 Db 121 GTGAAGAATGCTTGTGTTGTCTGTGAGATATGATGCGCTAAATAAGTAAACAACAGC 180



Matches	572:	Conservative	0:	Mismatches	5:	Indels	1:	Gaps	1:
OY	2687	atgaaatcacaatcattagaaacaataatcagctatgctatgcatctatcaacagatgaatgct	2746						
DB	1	ATTAAATCATCATCAATTAAACAACAATATCAGCTATGTGCAATTACCTATCAACAGATGAATGCT	60						
OY	2747	atcaagtaagcaagtgctttggccgaagaacttcaaaagccttcccggttaagcgttc	2806						
DB	61	ATCAAGTAAGCAACAATGTTTGGCCAGAAACATCCAAAGGCGCTTTCGCCGTTTAAACGGCTTC	120						
OY	2807	caactgagatataatgcaatctgctccctttgtgcacaaagatccgttaaagagagaag	2866						
DB	121	CACTTGAGTATATGCGCAATCTGTGCGCTTTGTGGCAAAAGATCTTTAAAGAGAGAAAGN	180						
OY	2867	ctcactgtaggcaatgctttgtgtgaaaaaalataaactgtaagcgcgagatctgtaagcagc	2926						
DB	181	CTCATGCTATNGCAATAGTTGGTGAATAATGTAAATGTAAAGCGGAGATATGTGAAGCAGC	240						
OY	2927	atgcagcgtctagtgaaaaaatatattgtctctctctcaccagagatattgttccatatcaa	2986						
DB	241	ATGCAAGCTGTAAGTAAATAATATGTCCTCTCCACAGAGATGTTGTTCCATATTACAA	300						
OY	2987	ttcaccttttgacacatgaccacagatattgtcaaaagtacagatatgaacaacttaag	3046						
DB	301	TTCACCTTTTGGCACAATACCCAGATATATGTCAAAGTACAGAGATATTGAACACTTTAAG	360						
OY	3047	atgttaagaatgtcttctgtgttctgtctgcgtgaataataatggtctaaataatgaataaca	3106						
DB	361	ATGTTAAAGAAATGCTTGTGTTGTTGTGGAATAATTAATGGCTTAAATAATGAATAACA	420						
OY	3107	gtcacgctttatcaagaagaatggttagaataatataataaacaacaaagatgccaaagac	3166						
DB	421	GTCACGCTTTTATCATCGAAAGATGTGAGAAATATATTAACAACAAACAAAGATGCCAATGAC	480						
OY	3167	cagatgagtcacaaaatgtaataaacaacgtgcacgtgtgtgattgttcattgaatatca	3226						
DB	481	CAGATGATGCGAATAATGATGAATAAAGTGTACAGTGTGTGATGTGTCGATGAATATCA	540						
OY	3227	tcatgtcaaaagagtactacatcac-aggttgaatctcc	3263						
DB	541	TCATGTCAAAGAGTACTCATATACNAGTGTGGAATCTCC	578						
RESULT	8								
BI142489		944 bp	mRNA	EST	24-OCT-2000				
LOCUS	6017985573P1	NCI_CGAP_Lu30	Mus musculus	CDNA clone	IMAGE:4016155 5',				
DEFINITION	mRNA sequence.								
ACCESSION	BI142489								
VERSION	BI142489.1	GI:10981529							
KEYWORDS	EST.								
SCURCE	house mouse.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
TITLE	1 (bases 1 to 944)								
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .								
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)								
	Unpublished (1999)								
	Contact: Robert Strausberg, Ph.D.								
	Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a>								
	Tissue Procurement: Gilbert Smith, Ph.D.								
	CDNA Library Preparation: Life Technologies, Inc.								
	Genomic Library Arrayed by: The I.M.A.G.E. Consortium (LMU)								
	DNA Sequencing by: Incyte Genomics, Inc.								
	Clone distribution: NCI-CGAP clone distribution information can be								
	found through the I.M.A.G.E. Consortium/LMU at:								
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>								
	Plate: L1AM9263	row: a	column: 20						
	High quality sequence stop: 620.								
FEATURES	location/Qualifiers								
source	1..944								
	/organism="Mus musculus"								

[illegible]







```

QY 543 gaccctattatttgcgaaggatgatacagtgctcaggaagcttggatcgglttagt 602
    |||||||
DB 180 GACCTATATATTTGTAAGGTGATACAGTGTCTCAGGACCTTTGGATACGGTTTAGT 239
QY 603 aaatcgtgtaaccgtcgtcattaaagaatttaacaagcaagcatgatattggcaagcttt 662
    |||||||
DB 240 AAATTCGTAACCTGCTCATGAAGATTAAACAAGCAAGCAATGATTTGGCAAAAGGCTTT 299
QY 663 actgaagagagacagcctcaagctattgagcattatataccacttttatacaggtct 722
    |||||||
DB 300 ACTGAAGAGAGACAGCTCAAGCTATTGAGCATTATTACCAATTTTAACTGACGTTCT 359
QY 723 gatgcttgggaaacaactatcagcagattgtcagaagcagctc-ttgacttaatttgg 781
    |||||||
DB 360 GATGCTTGGGAAACAATCATCATCAGCATTTGTGAGAGACATGCTTTGACTTAATTTTGG 419
QY 782 agctcacaatatattgtagtattgctctcctcgttcttaccaccgc-ttgaatttaa 840
    |||||||
DB 420 AGCTTACATATATGTAAGTCAATTTGCTGCTCTGTTTAAACCCAGCTTTGAATTTAA 479
QY 841 ttaagaagcaatgataatgagagcagcctcaagcttcttaactactgcaaaatgctt 900
    |||||||
DB 480 TTAAGAAGCAATGATATATGAGAGAGCCCTTACAGTTGTTAACTACTGCAAAAATGTTT 539
QY 901 ggggcaagagatcagaattgctctc-aaaaaagcacttggcagtgctac-ttgg 958
    |||||||
DB 540 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAAACAAGCCACTTTGGCAGTGCTACTTTGG 599
QY 959 gaaggttaataatgatacattgatacctca-ccgcctggaagtgtga-aatttgctgcca 1016
    |||||||
DB 600 GCAAGTTTAAATGATATGCAATGACCAATTCGCTGGAATGTGACCAATTTGCTGACCA 659
QY 1017 ttgtctcagtaacacatccctgatttgc-aaagacttaacagagatcctt-aaagtgaag 1074
    |||||||
DB 660 TTGCTCATGAGAACCATTCGATTACCAAAAAGATTAAACAGAGTATCTTAAAGTGAGG 719
QY 1075 tcaatgacc 1084
    |||||
DB 720 TCACCTTGAAAC 729

```

RESULT 5

LOCUS BE504550 596 bp mRNA EST 04-AUG-2000

DEFINITION hz58a05.x1 NCI\_CGAP\_Lu24 Homo sapiens CDNA clone IMAGE:3212144 3' similar to TR:094237 O94237 HYPOTHEICAL 45.2 KD PROTEIN ;, mRNA sequence.

ACCESSION BE504550

VERSION BE504550.1 GI:9706958

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 596)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/HLN, send email to: [info@imgc.llnl.gov](mailto:info@imgc.llnl.gov)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 458.

FEATURES

Source 1..596 Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3212144"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 211 a 103 c 117 g 165 t
ORIGIN
Query Match 14.3%; Score 596; DB 136; Length 596;
Best local Similarity 100.0%; Pred. No. 1.9e-134;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2687 atgaatcatcacatcagaacatcagctatgctcattagctatcacaagatgct 2746
    |||||||
DB 1 ATGAATCATCACATTAAGAAATATCAGCTATGTGCTATGCTATCAACGATGAATGCT 60
QY 2747 atcaatgaagaagaagtgctggccagaagaactcacaagaagccttccgtttacgcttc 2806
    |||||||
DB 61 ATCAAGTAAACAAAGTGTGGCCAGAACTTCACAAAGGCTTTCCGTTACGGCTTC 120
QY 2807 cacttgaglatagcaatcgtgccccttgctgcaaaagaatccgtlaaagagagaag 2866
    |||||||
DB 121 CACTTGAGTATATGCAATCTGTCCTTTGTGCAAAAAGATCCGTGAAGAGAGAGAG 180
QY 2867 ctcatgctaggaatgtttggtgaaaaataataatgaagcggtgagctatcgaagcagc 2926
    |||||||
DB 181 CTCATGCTAGGCAATGTTTGGTGAATAATTAATGAAGCGGAGATATCTGAAGCAGC 240
QY 2927 atgcaactgtagtgaaaaattatgtctctcttaccagaagatgtgtgtccatatcaa 2986
    |||||||
DB 241 ATGCAGCTGTTATGAAAAATTTATGCTCTTTTACCAAGTATGTTGCTCATATACAA 300
QY 2987 ttcaacttttggacatgaccagatattgtcaagaagatcagatattgaacaaactaaag 3046
    |||||||
DB 301 TTTCACCTTTTGGCAATGACCCAGATATGCTCAAGATGAGATATGAACAACCTTAAG 360
QY 3047 atgttaagaagatgcttggcttggcttgcgaataataatgcttaaaatgaataaaca 3106
    |||||||
DB 361 ATGTTAAAGAAATGCTTTGCTTGTGGAATAATTAATGAGCTTAAATAATGAATAATACA 420
QY 3107 gtcaagctttatcagaagaatgtgtagaataatttaaaacaaacaaagaatgccaagaag 3166
    |||||||
DB 421 GTACAGCTTTTATCAGAAATGATGTAATAATTTTAAACAACAAAGATGCCAAGGAC 480
QY 3167 cagatgatgcacaaatgaatgaacaaactgtacactgtgtgtgtgtgtgtgtgtgtgtgt 3226
    |||||||
DB 481 CAGATGATGCACAAATGAATGAATAAATCTGTACACTGTGTGATGTGTGCATGATATCA 540
QY 3227 tcatgtcaagaagtaactacatcagtttggaaatcctctaaagaccggttaccaca 3282
    |||||||
DB 541 TCATGTCAAAAGAGTACTACATACATGATTGGAATCTCTTAAAGACCCGGTACTACCA 596

```

RESULT 6

LOCUS BG387646 802 bp mRNA EST 12-MAR-2001

DEFINITION 602412496f1 NIH\_MGC\_92 Homo sapiens CDNA clone IMAGE:4521095 5', mRNA sequence.

ACCESSION BG387646

VERSION BG387646.1 GI:13281092

KEYWORDS EST.

SOURCE human.



Tissue Procurement: DCTD/DTF/Gazdar  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: LLM9663 row: 9 column: 24  
 High quality sequence stop: 654.

## FEATURES

source  
 1. 848  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3886511"  
 /clone\_id="NIH\_MGC\_69"  
 /tissue\_type="large cell carcinoma, undifferentiated"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 261 a 163 c 191 g 232 t 1 others  
 ORIGIN

Query Match 14.9%; Score 621.2; DB 141; Length 848;

Best Local Similarity 93.1%; Pred. No. 1.5e-140; Indels 9; Gaps 6;

Matches 716; Conservative 0; Mismatches 44; Indels 9; Gaps 6;

2247 gccctgcataagagcctagatcccaagaacatgcgaacatcctacacacattgtgtac 2306

1 gccctgcataagagcctagatcccaagaacatgcgaacatcctacacacattgtgtac 60

2307 taattgtatattgtctctcccttgcaactgcgataatttgcctccttggaaattcttgg 2366

61 TA-TGGTATATTTGCTCTCTGACCTGATCAATTTGCTCTCTGAAATCTTTGGT 119

2367 agctactcttcttggaagatctctcatgaatgcgctcgaagggaagaagacaac 2426

120 AGCTACTTTCATTCGAAAGATCTTCTCATGATATCGGCTCCAGGAAAAAGCAAC 179

2427 taactcttggttccagatgaagaagatctccttgagacaatggtcacaattcaggtat 2486

180 TAAACTTGGGTTCCAGATGAAGATATCTCTGAGCAATGTCATAATTCAGGCTAT 239

2487 taaatgtgtgtcgtatgctctcttgcaactgcgataatttgcctccttggaaattcttgg 2546

240 TAAATATGATGCTTCATGCTTGAATGAATAATATCACTAATTCAGGAATTC 299

2547 taccttaagattgtcaacaacatatctcatagtgatggaagcttgacagagagga 2606

300 TACCTTAAGATTGCTAACACAATATTCATATGATGAGAGCTTGACAGAACAGGGA 359

2607 aattagtaaacagatatgtcaagctctgagactgtcgtggaagtgtctattgtgaagt 2666

360 AATTAGTAACACAGATATGTCTACGCTGAGACTTCTGCTGGAGTGTATGTGAAGCT 419

2667 ggcacaaagaaacctgttaccatgaataatcatcatagaacaatatcagctgtgcatc 2726

420 GGCACAAAGAACCTGTTCACATGAATATCATCATTAGAACATATACACTTGTGCAAT 479

2727 agctacaagatgaatgctatcaagtaagaagaagtgtttgccgaagaacctcacaaag 2786

480 AGCTATCAACGATGATGCTATCAAGTAAGCAAGTGTTCGCCGAAGAACTTCACAAAG 539

2787 ccttcccgcttac-ggtctcacttgatatagtcaatcgtgtcccttctgtcaa-aa 2844

540 CCTTCCCGCTTACGGGCTTCCACTTGATATATGCAATCTGTGCCCTTGGTGAAGAA 599

2845 gatcctgtaaa---ggagagaagaagctcaatgctgaagcaatgtttgtgaaaaataataat 2901

600 GATCCTGTAAANGAGAGAAAGAGCTCATGCTAGCAATGTTTGGTGAACATATTAAT 659

OY 2902 -gtaagcggagagatcatcgaagcagcatgcagctgttagtgaataattatgtctctc 2959

DB 660 TGTCAGGGGGGAGGATCTTGAAGCAATGTGGCGTTAGTAGAAGAACTTATGGTCTTC 719

OY 2960 taccagagatgtgttccatataacattcacccttggacacatgcc 3008

DB 720 TACCACAG-ATTTGGTGGCCATATGATGATCACTTTTGGGCGCTTGGCCC 767

## RESULT 4

BC432960 729 bp mRNA EST 14-MAR-2001

LOCUS 602495783P1 NIH\_MGC\_75 Homo sapiens CDNA clone IMAGE:4609533 5',

DEFINITION mRNA sequence.

ACCESSION BC432960

VERSION BC432960.1 GI:13339466

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/

AUTHORS NIH (bases 1 to 729)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LHCMI35 row: e column: 22

High quality sequence stop: 722.

Location/Qualifiers

1. 729

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4609533"

/clone\_id="NIH\_MGC\_75"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1:

5'-ATTCTAGAGGCCGAGCGCGGACATG-dT(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.65

kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH-MGC library."

BASE COUNT 225 a 129 c 144 g 231 t

ORIGIN

Query Match 14.7%; Score 612.4; DB 153; Length 729;

Best Local Similarity 97.9%; Pred. No. 2e-138;

Matches 715; Conservative 0; Mismatches 6; Indels 9; Gaps 9;

OY 363 gagccacacatcaatagattatttatttacttagaacattgcttggatcaatc 422

DB 1 GAGCCACAGCTTCATATGATATTTTATTACTTGAGAACATGCTGGGTCAAGTCATA 59

OY 423 taacatattgtttagttagaagatagcaatgaatttaccacagctatcacagaacct 482

DB 60 TAAATATGCTTGTGATGATGAGATAGCAATGAATTTTACCCAGCTATACAGAACTT 119

OY 483 atttcagttataaacaatgcccacaatcagaagaatcctatgacatggtgaccttat 542

DB 120 ATTTTCAGTTATTAACAATGAGCCACAATCAATCAGAAAGTCATATGATGATGAT 179



Matches	700: Conservative	0: Mismatches	21: Indels	4: Gaps	2:
3352	tcaaaagtactacatcacagtttggatctctcctaagaaccggtactaccagcttc	3291			
Db	1 TCAAGAGTACTACTACATACAG - TTGGAAATCTCTTAAGACCCGGTACTACAGCTCGTTTC	59			
3292	ttcactcaacctggaagaatttaagtaaacccaanaattatctgcctccctgaaatgaa	3351			
Db	60 TTCACTCAACTGCACACAAATTTTCAGTACACCAAAAATTTATCTGCTCTCTGAAATGAAA	119			
3352	tcaattttcactcctgynaaccttaaaaaccaaatglttctaagagctgttaacaagcca	3411			
Db	120 TCATTTTTCACCTCCTCGGAAAAACCTFAAACCAACCAATGTCTTAGAGAGCTGTTAACAAAGCCA	179			
3412	ctttcataagcaggaagaacaaatcttaagcacaatcatcacgaatggnaacgttaagcaat	3471			
Db	180 CTTTATATAGCAGGCAACCAATCTTAGACCAAAATATATACAAATGGAACGTGAAGCAT	239			
3472	gcaagcagcagctcacaatccaaagctctctctggaagaataaaaggggagagcttgatagttct	3531			
Db	240 GCAAGCAGCAGCTCAAAATCCAAAGCTCTCTGGAAGAAATAAAGGGGAGCTTGATAGTTCT	299			
3532	gaaatggaatcacaagtgaagaatgaagattacaacaatglttcaaccttgcgggggaaaaaa	3591			
Db	300 GAAATGGATCTACAGAGGAAATGAAATTCACCAATGTCTTCACCTTTCCGGGAAAAAA	359			
3592	agtgcaagagagagacagacttgcattctgttaaggtctgtaatttggaagaagcttagaagcag	3651			
Db	360 AGTGCAGAGAGAGACAGACTTGTATCTTTAAAGTCTGTAATTTGGACAACTTAGAGCGAGG	419			
3652	aaaaaaagcccggtcacagaacagagagagaataatgaatgatalgatatgacttgactaagttg	3711			
Db	420 AAAAAAAGCCCGCTCACAGAACAGAGGAGAAATAGATAGATGACTTTCATTAAGTTG	479			
3712	gtacaggaagcagaaaccttaaaagcgatcagaggaagtcgaaaaaagagcgatacggttca	3771			
Db	480 GTACAGGAGACGAAACCTTAAGGCGAGTACGGCAAGTCCGAAAGAGGCCATACGGCTTCA	539			
3772	gaatctgaatgaacagcaatggccttgaggaagaaagagctcaagaagaatataatagaanaa	3830			
Db	540 GAATCTGATGAACACACATGTGGCTGTAGGAAAAAGAGCGCTCAAAAGAAATATATTGAAACAA	599			
3831	--tgaagtgtaacgaatcagtcgcgcgcaaaaaaaggtaagaagagccgaccccaaaaacct	3888			
Db	600 CTGAACGATGAACCAAAATAGTCCGCCAAACCAAGGCTACAAAGAGCGCGACCAACAAACCT	659			
3889	cttggatgaaggtaaccacaagaagaagcacaacatgtaanaactcttaaaaaagaagacaa	3948			
Db	660 CTTGGTGGAGGTACACCAAAAGAAAGAGGCCAAGATGAAGAACTTGTATATAAAGAAAGCAAA	719			
3949	aaaaa 3953				
Db	720 CAAAA 724				
RESULT 3					
LOCUS	BE873840	848 bp	mRNA	EST	20-OCT-2000
DEFINITION	6014839J2F1 NIH_MGC_69 Homo sapiens CDNA clone IMAGE:3886511 5',				
ACCESSION	BE873840				
VERSION	BE873840.1	GI:10322616			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	NIH-MGC http://mgi.nci.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email:cgapbs-remail.nih.gov				



## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	716	17.2	791	BF509252	BF509252 UT-H-B14
2	663.4	15.9	745	BF509252	BF509252 UT-H-B14
3	621.2	14.9	848	BF509252	BF509252 UT-H-B14
4	612.4	14.7	729	BF509252	BF509252 UT-H-B14
5	596	14.3	596	BF509252	BF509252 UT-H-B14
6	591.4	14.2	802	BF509252	BF509252 UT-H-B14
7	559.2	13.4	578	BF509252	BF509252 UT-H-B14
8	532.8	12.8	944	BF509252	BF509252 UT-H-B14
9	503.8	12.1	608	BF509252	BF509252 UT-H-B14
10	495	11.9	683	BF509252	BF509252 UT-H-B14
11	482.6	11.6	535	BF509252	BF509252 UT-H-B14
12	478.4	11.5	480	BF509252	BF509252 UT-H-B14
13	477.4	11.4	608	BF509252	BF509252 UT-H-B14
14	474.8	11.4	490	BF509252	BF509252 UT-H-B14
15	468	11.2	888	BF509252	BF509252 UT-H-B14
16	466.8	11.2	470	BF509252	BF509252 UT-H-B14
17	423	10.1	423	BF509252	BF509252 UT-H-B14
18	406.8	9.7	648	BF509252	BF509252 UT-H-B14
19	392	9.4	434	BF509252	BF509252 UT-H-B14
20	388.4	9.3	391	BF509252	BF509252 UT-H-B14
21	379.2	9.1	735	BF509252	BF509252 UT-H-B14
22	377	9.0	445	BF509252	BF509252 UT-H-B14
23	376.4	9.0	738	BF509252	BF509252 UT-H-B14
24	368.2	8.8	595	BF509252	BF509252 UT-H-B14
25	365	8.7	703	BF509252	BF509252 UT-H-B14
26	360.2	8.6	437	BF509252	BF509252 UT-H-B14
27	360	8.6	486	BF509252	BF509252 UT-H-B14
28	353	8.5	580	BF509252	BF509252 UT-H-B14
29	352.4	8.4	885	BF509252	BF509252 UT-H-B14
30	351.8	8.4	429	BF509252	BF509252 UT-H-B14
31	331.8	8.0	896	BF509252	BF509252 UT-H-B14
32	329	7.9	408	BF509252	BF509252 UT-H-B14
33	326.8	7.8	701	BF509252	BF509252 UT-H-B14
34	326.4	7.8	403	BF509252	BF509252 UT-H-B14
35	321.6	7.7	902	BF509252	BF509252 UT-H-B14
36	320.2	7.7	421	BF509252	BF509252 UT-H-B14
37	320.2	7.7	504	BF509252	BF509252 UT-H-B14
38	315.2	7.6	560	BF509252	BF509252 UT-H-B14
39	311.4	7.5	738	BF509252	BF509252 UT-H-B14
40	310.6	7.4	662	BF509252	BF509252 UT-H-B14
41	309.4	7.4	883	BF509252	BF509252 UT-H-B14
42	302.6	7.3	751	BF509252	BF509252 UT-H-B14
43	298.6	7.2	570	BF509252	BF509252 UT-H-B14
44	298.6	7.2	777	BF509252	BF509252 UT-H-B14
45	297.6	7.1	580	BF509252	BF509252 UT-H-B14

## ALIGNMENTS

RESULT	1	791 bp	mRNA	EST	06-DEC-2000
BE509252	UT-H-B14-80W-C-07-0-UI-s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone				
LOCUS	IMAGE:3086220.3, mRNA sequence.				
ACCESSION	BE509252				
VERSION	BE509252.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLM at: www.bio.lnl.gov/bhrp/image/image.html  
Seq primer: M13 Forward  
POLA=yes.

## FEATURES

Location/Qualifiers  
1..791  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3086220"  
/clone\_lib="NCI\_CGAP\_Sub8"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT730-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; NCI\_CGAP\_Sub8 is a subtracted library derived from NCI\_CGAP\_Sub5. The NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI\_CGAP\_Sub4 (IMAGE clone ids 2723592-2729336; 25% of the driver population), NCI\_CGAP\_Sub6 (pool A1F-AJ0, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described [Bonald, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_LIB=NCI\_CGAP\_Lu5  
TAG\_TISSUE=lung  
TAG\_SEQ=CAAC

BASE COUNT 270 a 151 c 144 g 226 t  
ORIGIN

Query Match 17.2%; Score 716; DB 149; Length 791;  
Best Local Similarity 99.6%; Pred. No. 1,1e-163;  
Matches 749; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY	2680	tggtccatgaaatcattcaatgaacaaatcagctatgcatgcatcaagat	2739
DB	13	TCCTACATGAAATCATCATTAAGAACAAATATGAGCTATGCTATCAACGAT	72
QY	2740	gaatgcatgaaatgaagaagtggtgcccagaactcaagaagcttccggtta	2799
DB	73	GAATGCTATCAAGTAAAGACAGTGTTCGCCAGAACTTCAAGAGCTTCCGTTA	132
QY	2800	cggctccactgatatggaacatctgccccttggcacaagaatcccttaagaag	2859
DB	133	CGGCTTCCACTTGATATATGCAATCTGTGCTTTGGCAAAAGATCTTAAGAG	192
QY	2860	agaagaagctatgctaggaatgcttggtgaaataataatgtaaggcggagatctg	2919
DB	193	ACAAGAGCTCATGATGAGCAATGTTGTGGAATAATATTAAGCGGAGATATCG	252
QY	2920	aagcagcatgacgctgtagtgaataatatgctctctccacagagatgtgttcca	2979
DB	253	AAGCAGCATGACGCTGTATGTAATAATATTTCTCTTCAACAGATATTTGTCCA	312
QY	2980	tatacaatccacttggcacatgaccagatctgtcaagaagtaacagatatgaaca	3039
DB	313	TATACATTCACCTTTGGCACATGACCAATATATGTAAGATGAGATTTGACAA	372
QY	3040	cttaagaatgtaagaatgcttgggttggcttgcgaataataatggtctaaatgaa	3099



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 16:19:16 : Search time 6640.39 Seconds  
(without alignments)  
5940.428 Million cell updates/sec

Title:	US-09-512-581-3
Perfect score:	4173
Sequence:	I atgctcatctcaagactag.....cacaccacaaaaaaatgtg 4173

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.00

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

1:	Est1.*
2:	qb_est1.*
3:	qb_est2.*
4:	qb_est3.*
5:	qb_est4.*
6:	qb_est5.*
7:	qb_est6.*
8:	qb_est7.*
9:	qb_est8.*
10:	qb_est9.*
11:	qb_est10.*
12:	qb_est11.*
13:	qb_est12.*
14:	qb_est13.*
15:	qb_est14.*
16:	qb_est15.*
17:	qb_est16.*
18:	qb_est17.*
19:	qb_est18.*
20:	qb_est19.*
21:	qb_est20.*
22:	qb_est21.*
23:	qb_est22.*
24:	qb_est23.*
25:	qb_est24.*
26:	qb_est25.*
27:	qb_est26.*
28:	qb_est27.*
29:	qb_est28.*
30:	qb_est29.*
31:	qb_est30.*
32:	qb_est31.*
33:	qb_est32.*
34:	qb_est33.*
35:	qb_est34.*
36:	qb_est35.*
37:	qb_est36.*
38:	qb_est37.*
39:	qb_est38.*
40:	qb_est39.*
41:	qb_est40.*
42:	qb_est41.*
43:	qb_est42.*

14: em\_esthum11.\*  
 45: em\_esthum10.\*  
 46: em\_esthum12.\*  
 47: em\_esthum13.\*  
 48: em\_esthum14.\*  
 49: em\_esthum15.\*  
 50: em\_esthum16.\*  
 51: em\_esthum17.\*  
 52: em\_esthum18.\*  
 53: em\_esthum19.\*  
 54: em\_esthum20.\*  
 55: em\_esthum21.\*  
 56: em\_esthum22.\*  
 57: em\_esthum23.\*  
 58: em\_esthum24.\*  
 59: em\_esthum25.\*  
 60: em\_esthum26.\*  
 61: em\_esthum27.\*  
 62: em\_esthum28.\*  
 63: em\_estin1.\*  
 64: em\_estin2.\*  
 65: em\_estin3.\*  
 66: em\_estin4.\*  
 67: em\_estin5.\*  
 68: em\_eston2.\*  
 69: em\_estov1.\*  
 70: em\_estov2.\*  
 71: em\_estov3.\*  
 72: em\_estov4.\*  
 73: em\_estpl2.\*  
 74: em\_estpl3.\*  
 75: em\_estpl4.\*  
 76: em\_estpl5.\*  
 77: em\_estpl6.\*  
 78: em\_estpl7.\*  
 79: em\_estpl8.\*  
 80: em\_estpl9.\*  
 81: em\_estpl10.\*  
 82: em\_estpl11.\*  
 83: em\_estpl12.\*  
 84: em\_estpl13.\*  
 85: em\_estpl14.\*  
 86: em\_estpl15.\*  
 87: em\_estpl16.\*  
 88: em\_estpl17.\*  
 89: em\_estpl18.\*  
 90: em\_estpl19.\*  
 91: em\_estpl20.\*  
 92: em\_estpl21.\*  
 93: em\_estpl22.\*  
 94: em\_estpl23.\*  
 95: em\_estpl24.\*  
 96: em\_estpl25.\*  
 97: em\_estpl26.\*  
 98: em\_estpl27.\*  
 99: em\_estpl28.\*  
 100: em\_estpl29.\*  
 101: em\_estpl30.\*  
 102: em\_estpl31.\*  
 103: em\_estpl32.\*  
 104: em\_estpl33.\*  
 105: em\_estpl34.\*  
 106: em\_estpl35.\*  
 107: em\_estpl36.\*  
 108: em\_estpl37.\*  
 109: em\_estpl38.\*  
 110: em\_estpl39.\*  
 111: em\_estpl40.\*  
 112: em\_estpl41.\*  
 113: em\_estpl42.\*  
 114: em\_estpl43.\*  
 115: em\_estpl44.\*  
 116: em\_estpl45.\*



QY 1324 PEEEEEE 1331  
|||||

DB 551 PEEEEEE 558

## RESULT 50

SQ7720

lamin B - mouse

C:/Species: Mus musculus (house mouse)

C:/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 13-Aug-1999

C:/Accession: S07720; S12488

R:/Hoeger, T.H.; Krohne, G.; Franke, W.W.

Eur. J. Cell Biol. 47, 283-290, 1988

A:/Title: Amino acid sequence and molecular characterization of murine lamin B as deduced

A:/Reference number: S07720; MUID:89210899

A:/Accession: S07720

A:/Molecule type: mRNA

A:/Residues: 1-587 &lt;HOE&gt;

A:/Cross-references: EMBL:X16705

R:/Hoeger, T.H.

submitted to the EMBL Data Library, October 1989

A:/Reference number: S12488

A:/Accession: S12488

A:/Molecule type: mRNA

A:/Residues: 1-530, 'I', 532-544, 'T', 546-552, 'E', 553-587 &lt;HOE2&gt;

A:/Cross-references: EMBL:X16705; NID:g52868; PIDN:CAA34677.1; PID:g52869

C:/Superfamily: cytoskeletal keratin

C:/Keywords: nucleus

## Query Match

0.6%; Score 8; DB 2; Length 587;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
|||||

DB 552 PEEEEEE 559

Search completed: September 25, 2001, 10:06:54

Job time: 160 sec



A:Molecule type: DNA  
A:Residues: 1-568 <STO>  
A:Cross-references: GB:BA000008; NID:g8978870; PIDN:BAA98706.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: proS  
C:Superfamily: proline--trRNA ligase

Query Match 0.6%; Score 8; DB 2; Length 568;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 188 GDTVSOEL 195  
|||||  
DB 483 GDTVSOEL 490

RESULT 47  
F72070  
C:Proline--trRNA ligase (EC 6.1.1.15) - Chlamydomophila pneumoniae (strains CWL029 and AR39)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: F72070; C81598  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: F72070  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-568 <ARN>  
A:Cross-references: GB:AE001363; NID:g4376785; PIDN:RAD18640.1; PID:g437678  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: C81598  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-568 <REA>  
A:Cross-references: GB:AE002185; GB:AE002161; NID:g7189170; PIDN:AAF38117.1; PID:g718918  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: proS; CP0254  
C:Superfamily: proline--trRNA ligase  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 0.6%; Score 8; DB 2; Length 568;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 188 GDTVSOEL 195  
|||||  
DB 483 GDTVSOEL 490

RESULT 48  
A96766  
unknown protein F2p28 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A96766  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Andersen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: A96766  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-575 <STO>  
A:Cross-references: GB:AE005173; NID:g7109486; PIDN:AAF36750.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F2p9.28  
A:Map position: 1

Query Match 0.6%; Score 8; DB 2; Length 575;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

|||||

DB 188 EEEEEER 195

RESULT 49

VEHULB

Lamin B1 - human

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1991 #sequence\_revision 15-Nov-1996 #text\_change 10-Dec-1999

C:Accession: A34707; A57124

R:Pollard, K.M.; Chan, E.K.L.; Grant, B.J.; Sullivan, K.F.; Tan, E.M.; Glass, C.A.

Mol. Cell. Biol. 10, 2164-2175, 1990

A:Title: In vitro posttranslational modification of lamin B cloned from a human T-cell

A:Reference number: A34707; MUID:90220602

A:Accession: A34707

A:Molecule type: mRNA

A:Residues: 1-586 <POL>

A:Cross-references: GB:M34458; NID:g186877; PIDN:AAA36162.1; PID:g307106

R:Lin, F.; Worman, H.J.

Genomics 27, 230-236, 1995

A:Title: Structural organization of the human gene (LMNB1) encoding nuclear lamin B1.

A:Reference number: A57124; MUID:96044426

A:Accession: A57124

A:Status: translation not shown; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-586 <RES>

A:Cross-references: GB:L37747; NID:g576838; PIDN:NAC37575.1; PID:g576840

C:Comment: The lamins (A, B, and C) contains several alpha-helical domains capable of

rates with lamin dissociation; it does not reform until telophase, when the lamins ar

C:Genetics:

A:Gene: GDB:LMNB1

A:Cross-references: GDB:512284; OMIM:150340

A:Map position: 5q23.3-5q31.1

A:Introns: 120/2; 172/3; 214/3; 271/3; 313/3; 387/2; 462/3; 497/3; 537/3; 573/3

C:Function:

A:Description: structural component of the nuclear lamina, a fibrous meshwork on the

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; lipoprotein; membrane protein; methylated carboxyl end; nucl

F; 2-583/Product: lamin B1 #status predicted <MAT>

F; 2-34/Domain: head <HED>

F; 35-390/Domain: rod <ROD>

F; 35-71/Region: coil 1A

F; 82-240/Region: coil 1B

F; 244-390/Region: coil 2

F; 391-583/Domain: tail <END>

F; 415-418/Region: nuclear location signal

F; 583/Binding site: farnesyl (Cys) (covalent) #status predicted

F; 583/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predict

Query Match

Best Local Similarity 100.0%;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 8; DB 1; Length 586;



## RESULT 42

T04653  
receiver-like protein 4 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Nov-2000  
C:Accession: T04653; T52035  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15263  
A:Accession: T04653  
A:Molecule type: DNA  
A:Residues: 1-552 <BEV>  
A:Cross-references: EMBL:AL021636  
A:Experimental source: cultivar Columbia; BAC clone F10N7  
R:Juchholz, G.; Keitel, C.; Kircher, S.; Kudla, J.; Lohrmann, J.; Sweere, U.; Schaefer,  
submitted to the EMBL Data Library, May 1998  
A:Description: Nuclear-localized receiver-like proteins are differentially expressed in  
A:Reference number: Z25912  
A:Accession: T52035  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-552 <BUC>  
A:Cross-references: EMBL:AJ005195; PIDN:CAA06432.1  
C:Genetics:  
A:Gene: ARP4  
A:Map position: 4  
A:Introns: 43/1; 94/1; 207/1; 232/3; 517/3  
A:Note: F10N7.270  
C:Superfamily: response regulator homology  
F:19-129/domain: response regulator homology <RRH>

Query Match 0.6%; Score 8; DB 2; Length 552;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
|||||  
DB 164 EEEEEER 171

## RESULT 43

T49788  
related to merozoite surface antigen 2 [imported] - Neurospora crassa  
N:Alternate names: protein B9J10.250  
C:Species: Neurospora crassa  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49788  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49788  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-562 <SCH>  
A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.250  
A:Experimental source: BAC clone B9J10; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B9J10.250  
A:Map position: 6  
A:Introns: 386/2; 477/2; 527/1

Query Match 0.6%; Score 8; DB 2; Length 562;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
|||||  
DB 100 EEEEEER 107

## RESULT 44

S37241  
legumin B - fava bean  
C:Species: Vicia faba (fava bean)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S46503; S37241  
R:Heim, U.; Baumelein, H.; Mobus, U.  
Plant Mol. Biol. 25, 131-135, 1994  
A:Title: The legumin gene family: a reconstructed Vicia faba legumin gene encoding a  
A:Reference number: S46503; MUID:94272010  
A:Accession: S46503  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-564 <HE2>  
A:Cross-references: EMBL:Z26489; NID:g403335; PIDN:CAA81262.1; PID:g403336  
C:Genetics:  
A:Introns: 178/3; 439/3  
C:Superfamily: glycinin

Query Match 0.6%; Score 8; DB 2; Length 564;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
|||||  
DB 286 EEEEEER 293

## RESULT 45

A72329  
general secretion pathway protein E - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: A72329  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.N.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: A72329  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-566 <ARN>  
A:Cross-references: GB:AE001750; GB:AE000512; NID:g4981356; PIDN:AAD35919.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0837  
C:Superfamily: secretion protein xcpR

Query Match 0.6%; Score 8; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
|||||  
DB 167 EEEEEER 174

## RESULT 46

H86552  
prolyl tRNA synthetase [imported] - Chlamydomophila pneumoniae (strain J138)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H86552  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: H86552  
A>Status: preliminary



A:Reference number: Z14734; MUID:97305956

A:Accession: T10830

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-513 <PRE>

A:Cross-references: EMBL:AE000105; NID:g2182706; PID:g2182709

C:Genetics:

A:Gene: nlfk2

A:Genome: plasmid pNGR234a

C:Superfamily: dinitrogenase beta chain

C:Keywords: ATP; iron-sulfur protein; metalloprotein; molybdenum; nitrogen fixation; ox

Query Match 0.6%; Score 8; DB 2; Length 513;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 LRLTTTIL 858

|||||

DB 481 LRLTTTIL 488

RESULT 38

T00828

C:Species: Arabidopsis thaliana

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Mar-1999

C:Accession: T00828

R:de la Bastide, M.; Hameed, A.; Gnoj, L.; Jensen, K.; Shohdy, N.; Gottesman, T.; Haberm

McCombie, W.R.

submitted to the EMBL Data Library, January 1999

A:Description: A. thaliana BAC T13L16 from chromosome IV, top arm.

A:Reference number: Z14205

A:Accession: T00828

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-513 <DEL>

A:Cross-references: EMBL:AC003952; NID:g2708736; PID:g2708738

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Introns: 51/1; 57/1; 255/1; 283/3; 387/2; 411/2

A>Note: T13L16.2

Query Match

Best Local Similarity 0.6%; Score 8; DB 2; Length 513;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

|||||

DB 159 EEEEEER 166

RESULT 39

A56110

C:Species: Homo sapiens

C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 05-Nov-1999

C:Accession: A56110

R:Jackman, J.K.; Motto, D.G.; Sun, Q.; Tanemoto, M.; Turck, C.W.; Peltz, G.A.; Koretzky,

J. Biol. Chem. 270, 7029-7032, 1995

A:Title: Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein associated with c

A:Reference number: A56110; MUID:95221345

A:Accession: A56110

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-533 <JAC>

A:Cross-references: GB:U20158; NID:g806765; PIDN:AAC50135.1; PID:g806766

C:Genetics:

A:Gene: GDB:LCP2; SLP-76

A:Cross-references: GDB:1230199; OMIM:601603

A:Map position: 5q33.1-5qter

C:Superfamily: SH2 homology  
C:Keywords: phosphoprotein  
F:422-520/Domain: SH2 homology <SH2>

Query Match 0.6%; Score 8; DB 2; Length 533;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 SSPLPGKK 1197

|||||

DB 297 SSPLPGKK 304

RESULT 40

I36911

C:Species: Aotus trivirgatus (douroucoul

C:Date: 16-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 22-Jun-1999

C:Accession: I36911

R:Tseng, H.; Green, H.

Mol. Biol. Evol. 6, 460-468, 1989

A:Title: The involucrin gene of the owl monkey: origin of the early region.

A:Reference number: I36911; MUID:90014142

A:Accession: I36911

A:Molecule type: DNA

A:Residues: 1-544 <RES>

A:Cross-references: GB:M25313; NID:gl76557; PIDN:AAA35375.1; PID:gl76558

C:Comment: During the terminal differentiation of keratinocytes, this protein from th

linked envelope under the plasma membrane.

C:Superfamily: involucrin

C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

F:153-501/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

Query Match

Best Local Similarity 0.6%; Score 8; DB 1; Length 544;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 SQELDTV 199

|||||

DB 16 SQELDTV 23

RESULT 41

B44841

C:Species: Xenopus laevis (African clawed frog)

C:Date: 31-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 12-Apr-1995

C:Accession: B44841

R:Charnas, L.R.; Szaro, B.G.; Gainer, H.

J. Neurosci. 12, 3010-3024, 1992

A:Title: Identification and developmental

A:Reference number: A44841; MUID:92356194

A:Accession: B44841

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-544 <CHA>

A:Experimental source: brain

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIN:110225, NCBIP:110226)

C:Superfamily: cytoskeletal keratin

Query Match

Best Local Similarity 0.6%; Score 8; DB 2; Length 544;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1330

|||||

DB 462 APEEEEE 469



Hypertension 14, 435-444, 1989  
 A>Title: Molecular cloning of chromogranin A from rat pheochromocytoma cells.  
 A:Reference number: A60746; MUID:90007662  
 A:Accession: A60746  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-462 <P>  
 A:CROSS-references: GB:AF145445; NID:g5163367; PIDN:AA040652.1; PID:g5163368  
 R:Iacangelo, A.; Okayama, H.; Eiden, L.E.  
 FEBS Lett. 227, 115-121, 1988  
 A>Title: Primary structure of rat chromogranin A and distribution of its mRNA.  
 A:Reference number: S00291; MUID:88112232  
 A:Accession: S00291  
 A:Molecule type: mRNA  
 A:Residues: 1-107, 'QQQQ', 108-462 <IAC>  
 A:CROSS-references: EMBL:X06832; NID:g55950; PIDN:CAA29988.1; PID:g55951  
 R:Hutton, J.C.; Nielsen, E.; Kastern, W.  
 FEBS Lett. 236, 269-274, 1988  
 A>Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and F  
 A:Reference number: S02543; MUID:88312980  
 A:Accession: S02543  
 A:Molecule type: mRNA  
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Residues: 19-107, 'QQQQ', 108-462 <HUT>  
 A>Note: part of this sequence, including the amino end of beta-granin, was determined by  
 C:Superfamily: chromogranin A  
 C:Keywords: glycoprotein; phosphoprotein  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-462/Product: chromogranin A #status predicted <MAT>  
 F:92-107/Region: glutamine-rich  
 F:417-419/Region: cell attachment (R-G-D) motif  
 F:35-56/Disulfide bonds: #status predicted  
 F:68,147,216,349,358,386,408,409,413/Binding site: phosphate (Ser) (covalent) #status pr  
 F:189/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.6%; Score 8; DB 1; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1325 EEEEEER 1332  
 DB 340 EEEEEER 347  
 |||||

RESULT 34  
 A39868  
 chromogranin A precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A39868  
 R:Wu, H.J.; Rozansky, D.J.; Farmer, R.J.; Gill, B.M.; O'Connor, D.T.  
 J. Biol. Chem. 266, 13130-13134, 1991  
 A>Title: Structure and function of the chromogranin A gene. Clues to evolution and tissu  
 A:Reference number: A39868; MUID:91302337  
 A:Accession: A39868  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-463 <WUA>  
 A:CROSS-references: GB:M64278; NID:g192723; PIDN:AAA37457.1; PID:g192724  
 C:Superfamily: chromogranin A  
 F:92-116/Region: glutamine-rich

Query Match 0.6%; Score 8; DB 1; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1325 EEEEEER 1332  
 DB 341 EEEEEER 348  
 |||||

RESULT 35  
 G71231  
 probable replication factor C subunit - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000  
 C:Accession: G71231  
 R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu  
 DNA Res. 5, 55-76, 1998  
 A>Title: Complete sequence and gene organization of the genome of a hyper-thermo  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: G71231  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-468 <KAW>  
 A:CROSS-references: GB:AP000001; NID:g3236128; PIDN:BAA29182.1; PID:g3256499  
 A:Experimental source: strain OT3  
 A>Note: this accession replaces an interim accession for a sequence replaced by GenBa  
 C:Genetics:  
 A:Gene: PH0113  
 C:Superfamily: replication factor C large chain

Query Match 0.6%; Score 8; DB 2; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1325 EEEEEER 1332  
 DB 436 EEEEEER 443  
 |||||

RESULT 36  
 F84558  
 hypothetical protein At2g17970 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: F84558  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Talbot,  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: F84558  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-507 <STO>  
 A:CROSS-references: GB:AE002093; NID:g4406821; PIDN:AA020129.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g17970  
 A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1325 EEEEEER 1332  
 DB 153 EEEEEER 160  
 |||||

RESULT 37  
 T10830  
 nitrogenase (EC 1.18.6.1) molybdenum-iron protein beta chain - Rhizobium sp. (strain  
 C:Species: Rhizobium sp.  
 A:Variety: strain NGR234  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 19-Jan-2001  
 C:Accession: T10830  
 R:Freiberg, C.; Fellay, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A.; Perret, M.  
 Nature 387, 394-401, 1997  
 A>Title: Molecular basis of symbiosis between Rhizobium and legumes.



A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625  
 A:Accession: H86683  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-437 <STO>  
 A:Cross-references: GB:AE005176; NID:g12723351; PIDN:AAK04570.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: pil37

Query Match 0.6%; Score 8; DB 2; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1200 KRDSDLV 1207  
 |||||  
 DB 73 KRDSDLV 80

# RESULT 30

E70046  
 Iron transport system homolog yvrA - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001  
 C:Accession: E70046  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y.; M.; Odawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: E70046  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-442 <KUN>  
 A:Cross-references: GB:Z99120; GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15321.1; PI

A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yvrA  
 C:Superfamily: Bacillus subtilis probable iron transport system yvrA; ATP-binding casset  
 C:Keywords: ATP; nucleotide binding; P-loop  
 F:16-214/Domain: ATP-binding cassette homology <ABC>  
 F:33-40/Region: nucleotide-binding motif A (P-loop)

Query Match 0.6%; Score 8; DB 1; Length 442;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 KDLDLIK 514  
 |||||  
 DB 174 KDLDLIK 181

# RESULT 31

A37386  
 dnaB protein homolog - Chlamydia trachomatis (serotype D) plasmid pCHL1  
 C:Species: Chlamydia trachomatis  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
 C:Accession: A37386  
 R:Comanducci, M.; Ricci, S.; Cevenini, R.; Ratti, G.

Plasmid 23, 149-154, 1990  
 A:Title: Diversity of the Chlamydia trachomatis common plasmid in biovars with differ  
 A:Reference number: A37386; MUID:90301796  
 A:Accession: A37386  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-451 <COM>  
 A:Cross-references: GB:J03321; NID:g144462; PIDN:AAA91569.1; PID:g144465  
 C:Genetics:  
 A:Genome: plasmid  
 C:Superfamily: phage P22 gene 12 protein

Query Match 0.6%; Score 8; DB 2; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1169 RIKGRDLS 1176  
 |||||  
 DB 145 RIKGRDLS 152

# RESULT 32

S01921  
 dnaB protein homolog - Chlamydia trachomatis plasmids  
 C:Species: Chlamydia trachomatis  
 C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 15-Oct-1999  
 C:Accession: S01921; S01023; S01022  
 R:Comanducci, M.; Ricci, S.; Ratti, G.  
 Mol. Microbiol. 2, 531-538, 1988  
 A:Title: The structure of a plasmid of Chlamydia trachomatis believed to be required  
 A:Reference number: S01920; MUID:89013895  
 A:Accession: S01921  
 A:Molecule type: DNA  
 A:Residues: 1-451 <COM>  
 A:Cross-references: EMBL:X07547; NID:g40730; PIDN:CAA30421.1; PID:g40732  
 A:Experimental source: plasmid pCHL1  
 R:Hatt, C.; Ward, M.E.; Clarke, I.N.  
 Nucleic Acids Res. 16, 4053-4067, 1988  
 A:Title: Analysis of the entire nucleotide sequence of the cryptic plasmid of Chlamyd  
 A:Reference number: S00788; MUID:88233998  
 A:Accession: S01023  
 A:Molecule type: DNA  
 A:Residues: 88-451 <HAT>  
 A:Cross-references: EMBL:X06707  
 A:Experimental source: plasmid pLGV440  
 A:Accession: S01022  
 A:Molecule type: DNA  
 A:Residues: 1-78, 'IWTNRIS', 86 <HA2>  
 A:Cross-references: EMBL:X06707  
 A:Experimental source: plasmid pLGV440  
 C:Genetics:  
 A:Genome: plasmid  
 C:Superfamily: phage P22 gene 12 protein

Query Match 0.6%; Score 8; DB 2; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1169 RIKGRDLS 1176  
 |||||  
 DB 145 RIKGRDLS 152

# RESULT 33

A60746  
 chromogranin A precursor - rat  
 N:Contains: beta-granin; pancreastatin  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: A60746; S00291; S02543  
 R:Parmer, R.J.; Koop, A.H.; Handa, M.T.; O'Connor, D.T.



A:Reference number: A69250; MUID:98049343  
A:Accession: B69436  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-339 <KLE>  
A:Cross-references: GB:AF000999; GB:AF000782; NID:g2689322; PIDN:AAB89749.1; PID:g264906  
C:Superfamily: rat acidic ribosomal protein P0

Query Match 0.6%; Score 8; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331  
Db 317 PEEEEEE 324

RESULT 25  
G72775  
hypothetical protein APE0193 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: G72775  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A:Reference number: A72450; MUID:99310339  
A:Accession: G72775  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79105.1; PID:d1042881; PID:g5103388  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0193

Query Match 0.6%; Score 8; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 981 EKLSSLLP 988  
Db 325 EKLSSLLP 332

RESULT 26  
B71604  
rRNA methylase (SpoU family) (OO, TP) PFB0855c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71604  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743  
A:Accession: B71604  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-351 <GAR>  
A:Cross-references: GB:AF001421; GB:AF001362; NID:g3845293; PIDN:AAC71960.1; PID:g3845293  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0855c

Query Match 0.6%; Score 8; DB 2; Length 351;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1270 LKEDILEN 1277  
Db 305 LKEDILEN 312

RESULT 27  
S06324  
dnaB protein homolog - Chlamydia trachomatis plasmid pCTT1  
C:Species: Chlamydia trachomatis  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 11-Jan-2000  
C:Accession: S06324  
R:Stripakash, K.S.; MacAvoy, E.S. Nucleic Acids Res. 15, 10596, 1987  
A:Title: A gene for dnaB like protein in chlamydial plasmid.  
A:Reference number: S06324; MUID:88096599  
A:Accession: S06324  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-398 <SRI>  
C:Genetics:  
A:Genome: plasmid  
C:Superfamily: phage P22 gene 12 protein  
C:Keywords: DNA binding

Query Match 0.6%; Score 8; DB 2; Length 398;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1169 RIKGRLDS 1176  
Db 145 RIKGRLDS 152

RESULT 28  
T15827  
hypothetical protein C53C9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C:Accession: T15827  
R:Bentley, D. submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid C53C9.  
A:Reference number: Z18413  
A:Accession: T15827  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-418 <BEN>  
A:Cross-references: EMBL:U28734; NID:g861255; PID:g861256; PIDN:AAB52603.1; GSPDB:GNO  
A:Experimental source: strain Bristol N2; clone C53C9  
C:Genetics:  
A:Gene: CESP:C53C9.2  
A:Map position: X  
A:Introns: 25/3; 187/1; 310/2; 380/3; 405/3

Query Match 0.6%; Score 8; DB 2; Length 418;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331  
Db 369 PEEEEEE 376

RESULT 29  
H86683  
prophage p11 protein 37, capsid protein [imported] - Lactococcus lactis subsp. lactis  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H86683  
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmgren, K.; Weissenbach, J.; Eh Genome Res. in press, 2001



QY 1324 PEEEEEE 1331  
 |||||  
 DB 226 PEEEEEE 233

## RESULT 20

F84923  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: F84923  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: F84923  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-313 <STO>  
 A:Cross-References: GB:AE002093; NID:g5729705; PIDN:AAD48512.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g48120  
 A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 |||||  
 DB 80 EEEEEER 87

## RESULT 21

S58719  
 A:Alternate names: hypothetical protein YNL058c - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 16-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 21-Jul-2000  
 C:Accession: S58719; S62986  
 R:Bergez, P.; Doignon, F.; Crouzet, M.  
 Yeast 11, 967-974, 1995  
 A>Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV  
 A:Reference number: S58711; MUID:96021608  
 A:Accession: S58719  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-316 <BER>  
 A:Cross-References: EMBL:U12141; NID:gl1314216; PIDN:AAA99653.1; PID:g994828  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
 R:Bergez, P.; Doignon, F.; Crouzet, M.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S62975  
 A:Accession: S62986  
 A:Molecule type: DNA  
 A:Residues: 1-316 <BEW>  
 A:Cross-References: EMBL:Z71334; NID:gl1301928; PID:gl1301929; MIPS:YNL058C  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Map position: 14L  
 C:Keywords: transmembrane protein  
 F:95-111/Domain: transmembrane #status predicted <TM>

Query Match 0.6%; Score 8; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 |||||

DB 237 EEEEEER 244

## RESULT 22

JC6146  
 C:Card protein - Myxococcus xanthus  
 C:Species: Myxococcus xanthus  
 C:Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 08-Oct-1999  
 C:Accession: JC6146  
 R:Nicolas, F.J.; Cayuela, M.L.; Martinez-Argudo, I.M.; Ruiz-Vazquez, R.M.; Murillo, F.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 6881-6885, 1996  
 A>Title: High mobility group I(Y)-like DNA-binding domains on a bacterial transcript  
 A:Reference number: JC6146; MUID:96293442  
 A:Accession: JC6146  
 A:Molecule type: DNA  
 A:Residues: 1-316 <NIC>  
 A:Cross-References: EMBL:Z56280; NID:gl022325; PIDN:CAA91224.1; PID:gl022328  
 C:Comment: This protein is an eukaryotic transcription factor that is involved in bot  
 C:Genetics:  
 A:Gene: card  
 C:Keywords: leucine zipper; transcription factor

Query Match 0.6%; Score 8; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 KGRPPKP 1296

|||||

DB 257 KGRPPKP 264

## RESULT 23

Ti8283  
 A:hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
 C:Species: Dictyostelium discoideum  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: Ti8283  
 R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.  
 Genetics 148, 1117-1125, 1998  
 A>Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Dd  
 A:Reference number: Z14684; MUID:98198836  
 A:Accession: Ti8283  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-325 <RIE>  
 A:Cross-References: EMBL:U00796; NID:g2702254; PID:g2702258; PIDN:AAC18634.1  
 C:Genetics:  
 A:Introns: 85/1

Query Match 0.6%; Score 8; DB 2; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||

DB 159 PEEEEEE 166

## RESULT 24

B69436  
 A:LSU ribosomal protein L10E (rpl10E) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 13-Aug-1999  
 C:Accession: B69436  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Smith,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch



A:Reference number: S13625; MUID:91160746  
 A:Accession: S13625  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-233 <AOK>  
 A:Cross-references: GB:D90251; NID:g3135096; PIDN:BAA14291.1; PID:g2220291  
 A:Accession: S18497  
 A:Molecule type: protein  
 A:Residues: 115-146,'X',148-157,'X',159-162 <AOK1>  
 R:Hashimoto, Y.; Nagaoaka, I.; Yamashita, T.  
 Biochim. Biophys. Acta 1203, 236-242, 1993  
 A:Title: Purification of the antibacterial fragments of guinea-pig major basic protein.  
 A:Reference number: S40491; MUID:94092714  
 A:Accession: S40491  
 A:Molecule type: protein  
 A:Residues: 115-233 <HAS>  
 C:Superfamily: eosinophil major basic protein precursor; C-type lectin homology  
 C:Keywords: antibiotic; chondroitin sulfate proteoglycan; cytotoxin; eosinophil; glycoprotein  
 F:11-16/Domain: signal sequence #status predicted <SIG>  
 F:17-114/Domain: propeptide #status predicted <PRO>  
 F:115-233/Product: eosinophil major basic protein 1 #status experimental <MAT>  
 F:116-231/Domain: C-type lectin homology <LCH>  
 F:24-25/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:67/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted  
 F:134-231,208-223/Disulfide bonds: #status predicted

Query Match 0.6%; Score 8; DB 1; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 DB 73 PEEEEEE 80  
 |||||

RESULT 17  
 A:33729  
 C:Species: Oncorhynchus keta (chum salmon)  
 C:Date: 20-Sep-1991 #sequence\_revision 20-Sep-1991 #text\_change 20-Jun-2000  
 A:Accession: A23729; B49762  
 R:Takayama, Y.; Rand-Weaver, M.; Kawauchi, H.; Ono, M.  
 Mol. Endocrinol. 5, 778-786, 1991  
 A:Title: Gene structure of chum salmon somatolactin, a presumed pituitary hormone of the genus *Oncorhynchus*.  
 A:Reference number: A23729; MUID:92017883  
 A:Accession: A23729  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-233 <TA>  
 A:Cross-references: GB:D10638; NID:g222909; PIDN:BAA01485.1; PID:g222911  
 R:Takayama, Y.; Ono, M.; Rand-Weaver, M.; Kawauchi, H.  
 Gen. Comp. Endocrinol. 83, 366-374, 1991  
 A:Title: Greater conservation of somatolactin, a presumed pituitary hormone of the genus *Oncorhynchus*.  
 A:Reference number: A49762; MUID:92038679  
 A:Accession: B49762  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-233 <TA2>  
 A:Cross-references: GB:D10640; NID:g222903; PIDN:BAA01487.1; PID:g222904  
 C:Superfamily: prolactin

Query Match 0.6%; Score 8; DB 2; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 VSEKLLSL 986  
 DB 150 VSEKLLSL 157  
 |||||

RESULT 18

RHHUT  
 thyroliberin precursor [validated] - human  
 N:Alternate names: prothyrotropin; thyrotropin-releasing hormone precursor; TRH  
 N:Contains: prothyroliberin; thyroliberin  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jun-1990 #sequence\_revision 12-May-1995 #text\_change 08-Dec-2000  
 A:Accession: A34550  
 R:Yamada, M.; Radovick, S.; Wondisford, F.E.; Nakayama, Y.; Weintraub, B.D.; Wilber, M.L. Endocrinol. 4, 551-556, 1990  
 A:Title: Cloning and structure of human genomic DNA and hypothalamic cDNA encoding hu thyroliberin.  
 A:Reference number: A34550; MUID:91125361  
 A:Accession: A34550  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: GB:M63581; GB:M63582; NID:g190296; PIDN:AAA36480.1; PID:g190298;  
 C:Comment: The peptide hormone is released from the hypothalamus.  
 C:Genetics:  
 A:Gene: GDB:TRH  
 A:Cross-references: GDB:128072; OMIM:275120  
 A:Map position: 3pter-3qter  
 A:Introns: 71/1  
 A:Note: the first intron occurs before the initiator codon  
 C:Function:  
 A:Description: stimulates synthesis and secretion of thyrotropin from the anterior pituitary gland.  
 C:Superfamily: thyroliberin precursor  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; polypeptide; pyroglutamic acid  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-242/Product: prothyroliberin #status predicted <PTL>  
 F:84-86/Product: thyroliberin #status experimental <TL1>  
 F:114-116/Product: thyroliberin #status experimental <TL2>  
 F:135-137/Product: thyroliberin #status experimental <TL3>  
 F:152-154/Product: thyroliberin #status experimental <TL4>  
 F:186-188/Product: thyroliberin #status experimental <TL5>  
 F:227-229/Product: thyroliberin #status experimental <TL6>  
 F:84,114,135,152,186,227/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form)  
 F:86,116,137,154,188,229/Modified site: amidated carboxyl end (Pro) (amide in mature form)

Query Match 0.6%; Score 8; DB 1; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 DB 169 EEEEEER 176  
 |||||

RESULT 19  
 E84706  
 Hypothetical protein At2g30280 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: E84706  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Koo, H.; Moffat, R.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: E84706  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-294 <STO>  
 A:Cross-references: GB:AE002093; NID:g2347194; PIDN:AAC16933.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g30280  
 A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



A:Accession: B46315  
A:Molecule type: DNA  
A:Residues: 1-147 <DOO>  
A:Cross-references: GB:M28744; NID:g333136; PIDN:AAA47015.1; PID:g333137  
C:Superfamily: papillomavirus type 4 E4 protein  
C:Keywords: early protein

Query Match 0.6%; Score 8; DB 1; Length 147;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332  
|||||||

DB 102 EEEEEEE 109

## RESULT 12

hypothetical protein F22K18.210 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
C:Accession: T05575  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.  
A:Reference number: Z15419  
A:Cross-references: EMBL:AL035356  
A:Experimental source: cultivar Columbia; BAC clone F22K18  
C:Genetics:  
A:Map position: 4  
A:Introns: 85/3; 116/1; 169/3  
A:Note: F22K18.210  
C:Superfamily: Arabidopsis hypothetical protein F22K18.210

Query Match 0.6%; Score 8; DB 2; Length 204;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
|||||||

DB 60 PEEEEEE 67

## RESULT 13

JC1129  
nonhistone chromosomal protein HMG-2 - chicken  
N:Alternate names: high-mobility-group protein 2  
C:Species: Gallus gallus (chicken)  
C:Date: 05-Mar-1993 #sequence\_revision 05-Mar-1993 #text\_change 23-Jul-1999  
C:Accession: JC1129  
R:Sparrow, D.B.; Wells, J.R.E.  
Gene 114, 289-290, 1992  
A:Title: Sequence of a cDNA encoding chicken high-mobility-group protein-2.  
A:Reference number: JC1129; MUID:92290291

A:Accession: JC1129  
A:Molecule type: mRNA  
A:Residues: 1-207 <SPA>

A:Cross-references: GB:M80574; NID:g211928; PIDN:AAA48819.1; PID:g211929  
C:Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology  
C:Keywords: chromosomal protein; DNA binding; nucleus  
F:6-83/Domain: HMG box homology <HMG1>  
F:92-166/Domain: HMG box homology <HMG2>

Query Match 0.6%; Score 8; DB 2; Length 207;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

DB 187 PEEEEEE 194  
|||||||

## RESULT 14

JC1114  
high-mobility group protein 2 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 23-Jul-1999  
C:Accession: JC1114  
R:Davis, D.L.; Burch, J.B.E.  
Gene 113, 251-256, 1992  
A:Title: Isolation of a chicken HMG2 cDNA clone and evidence for an HMG2-specific 3'-  
A:Reference number: JC1114; MUID:92241676  
A:Accession: JC1114  
A:Molecule type: mRNA  
A:Residues: 1-207 <DAV>  
A:Cross-references: GB:M83235; NID:g211926; PIDN:AAA48818.1; PID:g211927  
C:Comment: The high mobility group proteins are among the most abundant nonhistone ch  
C:Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology  
C:Keywords: DNA binding; nucleus  
F:6-83/Domain: HMG box homology <HMG1>  
F:92-166/Domain: HMG box homology <HMG2>

Query Match 0.6%; Score 8; DB 2; Length 207;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
|||||||

DB 187 PEEEEEE 194

## RESULT 15

T49830  
hypothetical protein B24H17.160 [Imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49830  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49830  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-231 <SCH>  
A:Cross-references: EMBL:AL356815; GSPDB:GN00116; NCSP:B24H17.160  
A:Experimental source: BAC clone B24H17; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B24H17.160  
A:Map position: 6  
A:Introns: 38/1; 102/3; 128/3

Query Match 0.6%; Score 8; DB 2; Length 231;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332  
|||||||

DB 70 EEEEEEE 77

## RESULT 16

S13625  
eosinophil major basic protein 1 precursor - guinea pig  
C:Species: Cavia porcellus (guinea pig)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
C:Accession: S13625; S18497; S40491  
R:Aoki, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasa  
FEBS Lett. 279, 330-334, 1991  
A:Title: Sequencing and cloning of the cDNA of guinea pig eosinophil major basic prot



Query Match 0.6%; Score 8; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 EEEEEER 1332  
Db 19 EEEEEER 26

RESULT 8  
S27275  
GTP-binding regulatory protein gamma chain precursor - northern European squid  
N:Alternate names: G-protein gamma chain; transducin gamma chain  
C:Species: Loligo forbesi (northern European squid)  
C>Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 07-May-1999  
C:Accession: S27275; S17596; S25626  
R:Jott, J.S.; Ryba, N.J.P.; Pottinger, J.D.D.; Keen, J.N.; Carne, A.; Findlay, J.B.C.  
F:BS Lett. 312, 241-244, 1992  
A:Title: The gamma-subunit of the principal G-protein from squid (Loligo forbesi) photoreceptor  
A:Reference number: S27275; MUID:93050202  
A:Accession: S27275  
A:Molecule type: mRNA  
A:Residues: 1-87 <LOT>  
A:Cross-references: EMBL:Z15112; NID:g9509; PID:g9510  
A:Experimental source: retina  
R:Pottinger, J.D.D.; Ryba, N.J.P.; Keen, J.N.; Findlay, J.B.C.  
Biochem. J. 279, 323-326, 1991  
A:Title: The identification and purification of the heterotrimeric GTP-binding protein from squid (Loligo forbesi) photoreceptor  
A:Reference number: S17596; MUID:92028830  
A:Accession: S17596  
A:Molecule type: protein  
A:Residues: 'LE', '29', '71', '31-33', 'P', '35-41', '43-51', 'M', '53-59', 'D', '61-63', 'MK', '66', 'A', '68-70', 'X', 'Y', 'Z', 'AA', 'BB', 'CC', 'DD', 'EE', 'FF', 'GG', 'HH', 'II', 'JJ', 'KK', 'LL', 'MM', 'NN', 'OO', 'PP', 'QQ', 'RR', 'SS', 'TT', 'VV', 'WW', 'XX', 'YY', 'ZZ', 'AA', 'BB', 'CC', 'DD', 'EE', 'FF', 'GG', 'HH', 'II', 'JJ', 'KK', 'LL', 'MM', 'NN', 'OO', 'PP', 'QQ', 'RR', 'SS', 'TT', 'VV', 'WW', 'XX', 'YY', 'ZZ'  
F:85-87/Domain: heterotrimer; lipoprotein; membrane protein; prenylated cysteine; signal transduction  
F:84/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 0.6%; Score 8; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 EEEEEER 1332  
Db 12 EEEEEER 19

RESULT 9  
S20219  
ubiquinol--cytochrome-c reductase (EC 1.10.2.2) 11k protein precursor - human  
N:Alternate names: complex III 11k protein; cytochrome bcl complex 11k protein; mitochondria  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1988 #sequence\_revision 22-Jul-1994 #text\_change 21-Jul-2000  
C:Accession: S00219; I52694  
R:Ohta, S.; Goto, K.; Arai, H.; Kagawa, Y.  
F:BS Lett. 226, 171-175, 1987  
A:Title: An extremely acidic amino-terminal presequence of the precursor for the human mitochondrial cytochrome bcl complex 11k protein  
A:Reference number: S00219; MUID:88083627  
A:Accession: S00219  
A:Molecule type: mRNA  
A:Residues: 1-91 <OHT>  
A:Cross-references: EMBL:M36647; NID:g188564; PIDN:AAA36317.1; PID:g188565  
R:Jiliu, A.Y.; Bradner, R.C.  
Cancer Res. 53, 2460-2465, 1993  
A:Title: Elevated expression of the human mitochondrial hinge protein gene in cancer.  
A:Reference number: I52694; MUID:93265436  
A:Accession: I52694  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-56 <RES>  
A:Cross-references: GB:S61826; NID:g385936; PIDN:AAD13930.1; PID:g4261630  
C:Genetics:

A:Gene: GDB:UQCRH  
A:Cross-references: GDB:141852  
A:Map position: 22q13-22q13  
C:Superfamily: ubiquinol--cytochrome-c reductase 11k protein; ubiquinol--cytochrome-c  
C:Keywords: mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory chain  
F:1-13/Domain: transit peptide (mitochondrion) #status predicted <TRP>  
F:14-91/Product: ubiquinol--cytochrome-c reductase 11k protein #status predicted <MAT  
F:23-91/Domain: ubiquinol--cytochrome-c reductase 11k protein homology <U11>  
F:37-81.53-67/Disulfide bonds: #status predicted

Query Match 0.6%; Score 8; DB 1; Length 91;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331  
Db 16 PEEEEEE 23

RESULT 10  
A55819  
nonhistone chromosomal protein CHMG-1 - midge (Chironomus tentans)  
N:Alternate names: high mobility group protein I/Y homolog  
C:Species: Chironomus tentans  
C>Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 21-Jul-2000  
C:Accession: A55819; S47253; S47252  
R:Claus, P.; Schulze, E.; Wisniewski, J.R.  
J. Biol. Chem. 269, 33042-33048, 1994  
A:Title: Insect proteins homologous to mammalian high mobility group proteins I/Y (HM  
A:Reference number: A55819; MUID:95105193  
A:Accession: A55819  
A:Molecule type: mRNA  
A:Residues: 1-99 <CLA>  
A:Cross-references: EMBL:Z36898; NID:g534932; PIDN:CAA85365.1; PID:g534933  
A:Experimental source: clone pCM126  
R:Claus, P.; Schulze, E.; Wisniewski, J.R.  
submitted to the EMBL Data Library, August 1994  
A:Description: Insect proteins homologous to mammalian high mobility group proteins I  
A:Reference number: S47252  
A:Accession: S47253  
A:Molecule type: DNA  
A:Residues: 1-91, 'A', '93-99 <CL2>  
A:Cross-references: EMBL:Z36897; NID:g534886; PID:g534887  
A:Experimental source: clone pCM106  
C:Genetics:  
A:Introns: 15/3; 76/1  
A:Note: suggested by Southern blot analysis in Ref A55819 to be single copy gene  
C:Keywords: chromosomal protein; DNA binding  
F:8-13/Region: DNA-binding motif (K/R-G-R-G-R-P)  
F:55-60/Region: DNA-binding motif (K/R-G-R-G-R-P)  
F:75-80/Region: DNA-binding motif (K/R-G-R-G-R-P)

Query Match 0.6%; Score 8; DB 2; Length 99;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 KGRGRPSK 1380  
Db 55 KGRGRPSK 62

RESULT 11  
B46315  
E4 protein - human papillomavirus type 4  
C:Species: human papillomavirus type 4  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Jul-1999  
R:Doorbar, J.; Conerion, I.; Gallimore, P.H.  
C:Accession: B46315  
Virology 172, 51-62, 1989  
A:Title: Sequence divergence yet conserved physical characteristics among the E4 prot  
A:Reference number: A46315; MUID:89370332



## RESULT 5

S06005  
 N:Myosin alpha heavy chain, cardiac muscle [similarity] - rat  
 N:Alternate names: alpha-myosin heavy chain  
 N:Contains: myosin ATPase (EC 3.6.1.32)  
 C:Species: rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Jan-2001  
 C:Accession: S06005; S07535; A20971; A02988; I53305  
 R:McNally, E.M.; Gianola, K.M.; Leinwand, L.A.  
 Nucleic Acids Res. 17, 7527-7528, 1989  
 A:Title: Complete nucleotide sequence of full length cDNA for rat alpha cardiac myosin H  
 A:Reference number: S06005; MUID:90016822  
 A:Accession: S06005  
 A>Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1938 <MC>  
 A:Cross-references: EMBL:X15938; NID:956654; PIDN:CAA34064.1; PID:g56655  
 R:McNally, E.M.; Kraft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.  
 J. Mol. Biol. 210, 665-671, 1989  
 A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison  
 A:Reference number: S07535; MUID:90133919  
 A:Accession: S07535  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1938 <MC2>  
 R:Mahdavi, V.; Chambers, A.P.; Nadal-Ginard, B.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 2626-2630, 1984  
 A:Title: Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem.  
 A:Reference number: A20971; MUID:84194059  
 A:Accession: A20971  
 A:Molecule type: protein  
 A:Residues: 1-12, 'AP', 47-50, 'AP', 53-81, 'E', 83-86, 'Q', 88-109, 111-133, 'H', 135-16  
 R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.  
 Nature 297, 659-664, 1982  
 A:Title: Molecular characterization of two myosin heavy chain genes expressed in the adu  
 A:Reference number: A02988; MUID:82220036  
 A:Accession: A02988  
 A:Molecule type: mRNA  
 A:Residues: 1512-1574, 'S', 1576-1721, 'N', 1723-1851, 'N', 1853-1869, 'N', 1871-1933, 'I', 1935-1  
 A:Note: there are 10 or more myosin heavy chain genes in the rat, at least two of which  
 R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.  
 Eur. Heart J. 5, 181-191, 1984  
 A:Title: Cardiac myosin heavy chain isozymic transitions during development and under pa  
 A:Reference number: I53305; MUID:85179510  
 A:Accession: I53305  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1872-1933, 'I', 1935-1938 <RES>  
 A:Cross-references: GB:M32697; NID:g205596; PIDN:AAA41658.1; PID:g205597  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylate  
 F:87-767/Domain: myosin motor domain homology <MMOT>  
 F:177-184/Region: nucleotide-binding motif A (P-loop)  
 F:548-585/Region: actin binding #status predicted  
 F:656-678/Region: actin binding #status predicted  
 F:840-1938/Domain: coiled coil #status predicted <COI>  
 F:840-1280/Region: S2  
 F:1281-1938/Region: light meromyosin  
 F:128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
 F:183/Binding site: ATP (Lys) #status predicted  
 F:696,706/Active site: Cys #status predicted

Query Match 0.6%; Score 9; DB 1; Length 1938;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 557

557 EKIRKOLEV 565  
 |||||

## DB 1528

1528 EKIRKOLEV 1536  
 |||||

## RESULT 6

## I49464

alpha cardiac myosin heavy chain - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 02-Feb-2001  
 C:Accession: I49464; I49463; I49462; I49461; I49604  
 R:Quinn-Laquer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.  
 Genomics 13, 176-188, 1992  
 A:Title: Characterization of the allelic differences in the mouse cardiac alpha-myosi  
 A:Reference number: A38207; MUID:92250040  
 A:Accession: I49464  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1938 <RES>  
 A:Cross-references: GB:M76601; NID:gl91623; PIDN:AAA37162.1; PID:gl91624  
 A:Accession: I49463  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>  
 A:Cross-references: GB:M76600; NID:gl91621; PIDN:AAA37161.1; PID:gl91622  
 A:Accession: I49462  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1938 <RE3>  
 A:Cross-references: GB:M76599; NID:gl91619; PIDN:AAA37160.1; PID:gl91620  
 A:Accession: I49461  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-544, 'A', 546-1938 <RE4>  
 A:Cross-references: GB:M76598; NID:gl91617; PIDN:AAA37159.1; PID:gl91618  
 R:Gulick, J.; Subramaniam, A.; Neumann, J.; Robbins, J.  
 J. Biol. Chem. 266, 9180-9185, 1991  
 A:Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes  
 A:Reference number: I49604; MUID:91225025  
 A:Accession: I49604  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-67 <RE5>  
 A:Cross-references: GB:M62404; NID:gl92609; PIDN:AAA37424.1; PID:gl92610  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop  
 F:88-768/Domain: myosin motor domain homology <MMOT>  
 F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 0.6%; Score 9; DB 2; Length 1938;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 557

557 EKIRKOLEV 565  
 |||||

## DB 1529

1529 EKIRKOLEV 1537  
 |||||

## RESULT 7

T47289  
 hypothetical protein T14K23.20 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47289  
 R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.  
 Mayer, K.F.X.  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24458  
 A:Accession: T47289  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-81 <NVA>  
 A:Cross-references: EMBL:AL132909  
 A:Experimental source: cultivar Columbia; BAC clone T14K23  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 26/2; 59/1  
 A:Note: T14K23.20



979 944 ALCADPKVRRRAHARQCL 962  
980 468 ALCADPKVRRRAHARQCL 486  
981 944 ALCADPKVRRRAHARQCL 962  
982 468 ALCADPKVRRRAHARQCL 486  
983 944 ALCADPKVRRRAHARQCL 962  
984 468 ALCADPKVRRRAHARQCL 486  
985 944 ALCADPKVRRRAHARQCL 962  
986 468 ALCADPKVRRRAHARQCL 486  
987 944 ALCADPKVRRRAHARQCL 962  
988 468 ALCADPKVRRRAHARQCL 486  
989 944 ALCADPKVRRRAHARQCL 962  
990 468 ALCADPKVRRRAHARQCL 486  
991 944 ALCADPKVRRRAHARQCL 962  
992 468 ALCADPKVRRRAHARQCL 486  
993 944 ALCADPKVRRRAHARQCL 962  
994 468 ALCADPKVRRRAHARQCL 486  
995 944 ALCADPKVRRRAHARQCL 962  
996 468 ALCADPKVRRRAHARQCL 486  
997 944 ALCADPKVRRRAHARQCL 962  
998 468 ALCADPKVRRRAHARQCL 486  
999 944 ALCADPKVRRRAHARQCL 962  
1000 468 ALCADPKVRRRAHARQCL 486

## ALIGNMENTS

RESULT 1  
T00374  
hypothetical protein KIAA0648 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00374  
R:Fishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.  
DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A:Reference number: Z14142; MUID:98403880  
A:Accession: T00374  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-851 <ISH>  
A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0648

Query Match 1.4%; Score 19; DB 2; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 944 ALCADPKVRRRAHARQCL 962  
DB 468 ALCADPKVRRRAHARQCL 486  
RESULT 2  
A57573  
telomeric repeat binding factor - human  
C:Species: Homo sapiens (man)  
C>Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 19-May-2000  
C:Accession: A57573  
R:Chong, L.; van Steensel, B.; Broccoli, D.; Erdjument-Bromage, H.; Hanish, J.; Tempst,  
Science 270, 1663-1667, 1995  
A:Title: A human telomeric protein.  
A:Reference number: A57573; MUID:96099400  
A:Accession: A57573  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-439 <CHO>  
A:Cross-references: GB:U40705; NID:g2078442; PID:g1439542  
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo  
C:Keywords: DNA binding

F:375-428/Domain: myb DNA-binding repeat homology <MYB3>

Query Match 0.6%; Score 9; DB 2; Length 439;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

DB 53 APEEEEEEE 61

RESULT 3

T45840

GTPase activating-like protein - Arabidopsis thaliana

N:Alternate names: protein F2K15.210

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C:Accession: T45840

R:Kieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Benck

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23015

A:Accession: T45840

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-554 <RIE>

A:Cross-references: EMBL:AL132956

A:Experimental source: cultivar Columbia; BAC clone F2K15

C:Genetics:

A:Map position: 3

A:Introns: 53/2; 98/3; 115/2; 133/2

A:Note: F2K15.210

Query Match 0.6%; Score 9; DB 2; Length 554;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 SSSSNPSP 1167

DB 11 SSSSNPSP 19

RESULT 4

T49562

neurofilament-H related protein [imported] - Neurospora crassa

N:Alternate names: protein B208.100

C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49562

R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49562

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-837 <SCH>

A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.100

A:Experimental source: BAC clone B208; strain OR74A

C:Genetics:

A:Gene: NCSP:B208.100

A:Map position: 6

Query Match 0.6%; Score 9; DB 2; Length 837;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1287 KGKRGPRPK 1295

DB 695 KGKRGPRPK 703



833	7	0.5	1545	2	T46645	sulfonyleurea recep	906	6	0.4	39	2	D83721	hypothetical prote
834	7	0.5	1546	2	T42728	sulfonyleurea recep	907	6	0.4	40	2	G82271	hypothetical prote
835	7	0.5	1558	2	T29253	hypothetical prote	908	6	0.4	41	2	T14350	tubulin beta-3 cha
836	7	0.5	1560	2	T30282	calcium-binding pr	909	6	0.4	42	2	S78170	ribosomal protein
837	7	0.5	1567	2	T03730	antigen containing	910	6	0.4	42	2	T01991	hypothetical prote
838	7	0.5	1582	2	A56248	sulfonyleurea recep	911	6	0.4	42	2	T07285	hypothetical prote
839	7	0.5	1610	2	A46227	voltage-dependent	912	6	0.4	43	2	S02031	zinc-binding prote
840	7	0.5	1611	2	G84493	probable retroelem	913	6	0.4	44	2	C34327	tropinin T, fast-s
841	7	0.5	1613	2	S39059	protein BRG1 - hum	914	6	0.4	45	2	T52124	hypothetical prote
842	7	0.5	1638	2	T25352	hypothetical prote	915	6	0.4	47	2	S43651	proteasome subunit
843	7	0.5	1645	2	H85554	hypothetical prote	916	6	0.4	47	2	D40605	hmc 3'-region hypo
844	7	0.5	1646	2	JH0422	voltage-dependent	917	6	0.4	49	2	T12879	hypothetical prote
845	7	0.5	1647	2	S45252	SNF2beta protein -	918	6	0.4	50	2	B82702	hypothetical prote
846	7	0.5	1664	2	F84485	probable retroelem	919	6	0.4	51	2	T10865	hypothetical prote
847	7	0.5	1684	2	T02367	hypothetical prote	920	6	0.4	54	2	JS0853	hypothetical 5.8K
848	7	0.5	1706	2	I84499	zinc finger protei	921	6	0.4	54	2	T07284	hypothetical prote
849	7	0.5	1711	2	C71625	variant-specific s	922	6	0.4	55	2	T09464	H+-transporting AT
850	7	0.5	1717	2	T13961	DNA-directed RNA p	923	6	0.4	57	2	JN0740	hypothetical 6.7K
851	7	0.5	1720	2	T07258	cell division prot	924	6	0.4	59	2	C33556	prothymosin alpha
852	7	0.5	1721	1	I38902	retinoblastoma bin	925	6	0.4	60	1	HSAX22	histone H2B - Nile
853	7	0.5	1729	2	T18396	erythrocyte membra	926	6	0.4	61	2	S11646	hypothetical prote
854	7	0.5	1741	2	T13610	parallel sister ch	927	6	0.4	62	2	C71125	hypothetical prote
855	7	0.5	1753	2	T00350	hypothetical prote	928	6	0.4	63	1	H64095	carbon storage reg
856	7	0.5	1766	2	S03701	141k protein - pea	929	6	0.4	63	2	S07127	chymotrypsin/elast
857	7	0.5	1768	2	T13349	parallel sister ch	930	6	0.4	64	2	C64010	hypothetical prote
858	7	0.5	1774	2	B56101	collagen alpha 1(X	931	6	0.4	66	2	S61085	M protein precursor
859	7	0.5	1793	2	T47897	guanine nucleotide	932	6	0.4	66	2	H86678	hypothetical prote
860	7	0.5	1820	1	CHEE	sodium channel pro	933	6	0.4	67	2	S10555	glutathione transf
861	7	0.5	1824	2	T07589	disease resistance	934	6	0.4	69	2	A86808	hypothetical prote
862	7	0.5	1829	2	T34239	hypothetical prote	935	6	0.4	69	2	B85775	hypothetical prote
863	7	0.5	1829	2	T24583	hypothetical prote	936	6	0.4	69	2	C64925	hypothetical prote
864	7	0.5	1857	1	S01787	fatty-acid synthas	937	6	0.4	70	2	S73177	ribosomal protein
865	7	0.5	1857	2	T50513	hypothetical prote	938	6	0.4	70	2	A64011	hypothetical prote
866	7	0.5	1948	2	S00485	gene 11-1 protein	939	6	0.4	70	2	T23415	hypothetical prote
867	7	0.5	1966	2	T08991	hypothetical prote	940	6	0.4	70	2	T26824	hypothetical prote
868	7	0.5	2009	2	S49764	SEC7 protein - yea	941	6	0.4	70	2	T32870	hypothetical prote
869	7	0.5	2110	2	B41110	mycocosate synth	942	6	0.4	71	2	D42506	F-ORF-E protein -
870	7	0.5	2111	2	A70568	mycocosate synth	943	6	0.4	71	2	T32501	hypothetical prote
871	7	0.5	2149	2	C96695	ribulose biphosph	944	6	0.4	71	2	T29577	hypothetical prote
872	7	0.5	2161	2	JH0564	calcium channel al	945	6	0.4	72	2	S19237	histone H2B (clone
873	7	0.5	2181	2	A38198	calcium channel al	946	6	0.4	72	2	S78368	ribosomal protein
874	7	0.5	2203	2	T42742	voltage-dependent	947	6	0.4	72	2	C47072	marB protein - Esc
875	7	0.5	2206	2	G71611	hypothetical prote	948	6	0.4	72	2	B42645	ribosomal protein
876	7	0.5	2225	2	T26063	hypothetical prote	949	6	0.4	72	2	T47513	hypothetical prote
877	7	0.5	2265	2	T26183	hypothetical prote	950	6	0.4	72	2	G81564	ribosomal protein
878	7	0.5	2282	2	T42717	DNA-binding protei	951	6	0.4	73	2	S72246	dynelin heavy chain
879	7	0.5	2457	2	T18492	hypothetical prote	952	6	0.4	73	2	H71149	hypothetical prote
880	7	0.5	2924	2	T18378	variant-specific s	953	6	0.4	74	2	A05059	photosystem II pho
881	7	0.5	3026	2	T28431	variant surface pr	954	6	0.4	75	2	T03564	ribosomal protein
882	7	0.5	3135	2	A48584	transmission block	955	6	0.4	75	2	JQ2358	wheat aluminum ind
883	7	0.5	3147	2	T18674	hypothetical prote	956	6	0.4	76	2	S22203	photosystem I chal
884	7	0.5	3418	1	G02334	breast cancer tumo	957	6	0.4	76	2	S09897	hypothetical prote
885	7	0.5	4006	2	T29042	probable tenascin	958	6	0.4	76	2	T18091	hypothetical prote
886	7	0.5	4385	2	T29042	hypothetical prote	959	6	0.4	77	2	A47070	ferrochelatase (EC
887	7	0.5	4845	2	T31067	BiR repeat contain	960	6	0.4	78	2	C49039	T-cell receptor be
888	7	0.5	4859	2	S74173	ryanodine receptor	961	6	0.4	79	2	S76591	hypothetical prote
889	7	0.5	4859	2	S66572	ryanodine receptor	962	6	0.4	79	2	E64354	conserved hypotet
890	7	0.5	4872	2	S27272	ryanodine receptor	963	6	0.4	79	2	A86517	hypothetical prote
891	7	0.5	5035	1	I46646	ryanodine receptor	964	6	0.4	79	2	T35842	hypothetical prote
892	7	0.5	5037	1	A54161	ryanodine-binding	965	6	0.4	79	2	B72106	hypothetical prote
893	7	0.5	5037	2	B35041	ryanodine receptor	966	6	0.4	79	2	T13293	hypothetical prote
894	7	0.5	5170	2	T15348	hypothetical prote	967	6	0.4	81	2	C84077	nitrogen fixation
895	7	0.5	5175	2	T20992	hypothetical prote	968	6	0.4	81	2	T31007	conserved hypotet
896	7	0.5	5198	2	T43290	hemocentin precurs	969	6	0.4	81	2	D81565	hypothetical prote
897	6	0.4	12	2	S16335	beta-conglycinin a	970	6	0.4	81	2	F82745	hypothetical prote
898	6	0.4	18	2	A59137	protein P11 - gold	971	6	0.4	83	2	I46058	caldesmon - bovine
899	6	0.4	20	2	S19240	histone H2B (clone	972	6	0.4	83	2	H69175	hypothetical prote
900	6	0.4	20	2	A61093	glue protein - Cal	973	6	0.4	84	2	A57231	caldesmon - leech
901	6	0.4	22	2	B46236	transforming prote	974	6	0.4	84	2	T10345	hypothetical prote
902	6	0.4	22	2	S69351	N-methylhydantoin	975	6	0.4	87	2	S26963	hypothetical prote
903	6	0.4	30	2	PC2328	multicatalytic end	976	6	0.4	87	2	D82136	cell division topo
904	6	0.4	30	2	PD0014	cAMP response elem	977	6	0.4	87	2	JQ1683	UL49A protein - hu
905	6	0.4	37	2	G82917	ribosomal protein	978	6	0.4	88	1	QOEC8	hypothetical prote



687	7	0.5	886	2	A48586	suppressor of hair	760	7	0.5	1114	2	I50222	deltaErf1 - chicken
688	7	0.5	886	2	A54442	3',5'-cyclic-nucle	761	7	0.5	1115	2	B84476	probable TPR repea
689	7	0.5	888	2	T51593	GTP-binding regula	762	7	0.5	1116	2	T13854	nuclear protein SA
690	7	0.5	890	2	T00800	disease resistance	763	7	0.5	1117	2	JC4934	delta-crystallin/E
691	7	0.5	892	2	B46203	mating type A alph	764	7	0.5	1121	2	T25715	hypothetical prote
692	7	0.5	895	2	A45554	rhostry complex po	765	7	0.5	1124	2	JX0293	zinc finger protei
693	7	0.5	898	2	H84701	probable villin li	766	7	0.5	1137	2	S13759	morphogenesis-rela
694	7	0.5	899	2	T84477	probable retroelem	767	7	0.5	1141	2	A44093	CGMP-inhibited cAM
695	7	0.5	899	2	B48586	suppressor of hair	768	7	0.5	1148	2	A71446	mypl protein - smu
696	7	0.5	901	2	T01135	probable GTP-bind	769	7	0.5	1150	2	S98775	hypothetical prote
697	7	0.5	901	2	JC6093	dead ringer nuclea	770	7	0.5	1153	2	T00615	DNA polymerase III
698	7	0.5	902	2	G83635	probable CLPA/B-ty	771	7	0.5	1154	2	A56242	E-box-binding repr
699	7	0.5	906	2	T01440	hypothetical prote	772	7	0.5	1161	2	I59311	nardillysin (EC 3.4
700	7	0.5	907	2	T48500	Me12-like protein	773	7	0.5	1167	2	T13927	adenylate cyclase
701	7	0.5	909	2	S32538	CGMP-gated cation	774	7	0.5	1169	2	C84547	hypothetical prote
702	7	0.5	910	2	T50671	villin 1 (imported	775	7	0.5	1171	2	T13065	PI82 protein - fr
703	7	0.5	912	2	A54423	brevican precursor	776	7	0.5	1181	2	A43346	1-phosphatidylinos
704	7	0.5	913	2	T15474	hypothetical prote	777	7	0.5	1182	2	T30189	myelin transcripti
705	7	0.5	929	2	T28927	hypothetical prote	778	7	0.5	1184	2	A42904	adenylate cyclase ly
706	7	0.5	932	2	S62555	protoptast regene	779	7	0.5	1184	2	T41515	coiled coil protei
707	7	0.5	938	2	A56731	chromatin assembly	780	7	0.5	1187	2	T46637	transcription fact
708	7	0.5	943	2	B83068	probable oxidoredu	781	7	0.5	1188	2	T05324	hypothetical prote
709	7	0.5	944	2	G86720	translation initia	782	7	0.5	1188	2	T46608	zinc finger protei
710	7	0.5	946	2	T19465	hypothetical prote	783	7	0.5	1196	2	T13057	KIAA0729 protein -
711	7	0.5	949	2	D84487	probable retroelem	784	7	0.5	1197	2	T13956	timeless protein h
712	7	0.5	954	2	T24748	hypothetical prote	785	7	0.5	1197	2	D86317	protein F15H18.21
713	7	0.5	954	2	T22369	hypothetical prote	786	7	0.5	1199	2	T15826	hypothetical prote
714	7	0.5	957	2	G84528	hypothetical prote	787	7	0.5	1199	2	S20969	Na+/Ca2+, K+-exchan
715	7	0.5	960	2	T07680	VPS41 protein homo	788	7	0.5	1205	2	T13959	timeless protein T
716	7	0.5	960	2	T17297	hypothetical prote	789	7	0.5	1208	2	T00362	hypothetical prote
717	7	0.5	963	2	A5926	DNA binding protei	790	7	0.5	1210	2	A80001	phospholipase C (E
718	7	0.5	966	2	E84052	penicillin-binding	791	7	0.5	1214	2	T00356	hypothetical prote
719	7	0.5	967	1	HXAD2	hexon protein - hu	792	7	0.5	1217	2	T00270	hypothetical prote
720	7	0.5	968	2	T45746	hypothetical prote	793	7	0.5	1219	2	I61713	co-repressor prote
721	7	0.5	971	2	S23408	prematurely termin	794	7	0.5	1221	2	A49457	fibulin-2 precurs
722	7	0.5	972	2	T50400	origin recognition	795	7	0.5	1223	2	S29717	adenylate cyclase
723	7	0.5	972	2	T49773	related to actin-i	796	7	0.5	1227	2	B34911	band 3-related pro
724	7	0.5	973	2	T01862	hypothetical prote	797	7	0.5	1227	2	A33638	erythrocyte anion
725	7	0.5	975	1	A31497	kinesin heavy chai	798	7	0.5	1229	2	A56068	co-repressor prote
726	7	0.5	978	2	H86319	hypothetical prote	799	7	0.5	1231	1	A48490	endo-1,4-beta-xyla
727	7	0.5	979	2	A35913	regulatory factor	800	7	0.5	1232	2	I38496	anion exchanger 3
728	7	0.5	993	2	T17230	hypothetical prote	801	7	0.5	1232	2	T47993	hypothetical prote
729	7	0.5	999	2	S68689	glucose regulated	802	7	0.5	1233	2	S56271	hypothetical prote
730	7	0.5	1001	2	S30385	G9a protein - huma	803	7	0.5	1241	2	J00466	potassium transpor
731	7	0.5	1013	2	B96544	hypothetical prote	804	7	0.5	1262	2	T33074	hypothetical prote
732	7	0.5	1014	2	T31433	Na+/Ca2+, K+-exchan	805	7	0.5	1264	2	S41603	type V adenylyl cy
733	7	0.5	1022	2	I53078	homeotic gene regu	806	7	0.5	1268	2	T50252	probable transcrip
734	7	0.5	1026	2	B96663	hypothetical prote	807	7	0.5	1272	2	C96637	hypothetical prote
735	7	0.5	1032	2	T18293	quanylate kinase-i	808	7	0.5	1274	2	T04018	hypothetical prote
736	7	0.5	1034	2	S76134	hypothetical prote	809	7	0.5	1277	2	T32731	PAR interacting pr
737	7	0.5	1037	2	A60163	glycoprotein Iib -	810	7	0.5	1277	2	T14152	synaptic scaffold
738	7	0.5	1038	2	T02634	rep protein homolo	811	7	0.5	1280	2	T00365	hypothetical prote
739	7	0.5	1038	2	S52522	hypothetical prote	812	7	0.5	1283	2	T28812	hypothetical prote
740	7	0.5	1043	2	A56037	DNA-binding protei	813	7	0.5	1300	2	T03166	probable immediate
741	7	0.5	1045	2	T41119	internalin- relate	814	7	0.5	1360	2	T06699	zinc finger protei
742	7	0.5	1047	2	T21306	hypothetical prote	815	7	0.5	1361	2	S50943	hypothetical prote
743	7	0.5	1054	2	T14189	hypothetical prote	816	7	0.5	1386	2	T49316	profilaggrin relat
744	7	0.5	1058	2	T30580	p-type ATPase - sl	817	7	0.5	1388	2	T38720	chromodomain helic
745	7	0.5	1062	2	S61196	SUM1 protein - yea	818	7	0.5	1389	2	I50090	carboxypeptidase g
746	7	0.5	1063	2	D86731	hypothetical prote	819	7	0.5	1402	2	I46707	translation initia
747	7	0.5	1067	2	T28663	hypothetical prote	820	7	0.5	1428	2	T13926	probable protein p
748	7	0.5	1067	2	S35423	protein kinase sgg	821	7	0.5	1432	2	S58819	antiviral protein -
749	7	0.5	1068	2	G86452	protein AAR3127.1	822	7	0.5	1440	2	T27942	lin-15B protein -
750	7	0.5	1072	2	T50949	verprolin related	823	7	0.5	1443	2	I50600	neogenin - chicken
751	7	0.5	1075	2	T49879	hypothetical prote	824	7	0.5	1463	2	C86482	protein F5J5.1 [im
752	7	0.5	1078	2	T42712	myelin transcripti	825	7	0.5	1467	2	T48162	hypothetical prote
753	7	0.5	1087	2	T30330	geisolin-related p	826	7	0.5	1474	2	T20488	hypothetical prote
754	7	0.5	1094	2	S49313	protein kinase - s	827	7	0.5	1474	2	F69009	probable membrane
755	7	0.5	1097	2	S68685	adenylate cyclase	828	7	0.5	1479	2	T17401	transcription regu
756	7	0.5	1099	2	A55405	adenylate cyclase	829	7	0.5	1480	2	T05566	hypothetical prote
757	7	0.5	1100	2	T42260	quanylate cyclase	830	7	0.5	1493	2	A44224	DNA repair helicas
758	7	0.5	1108	2	A48508	cyclic-nucleotide	831	7	0.5	1511	2	T42711	sulfonylurea recep
759	7	0.5	1110	2	I51116	NF-180 - sea lamp	832	7	0.5	1545	2	T42751	sulfonylurea recep



541	7	0.5	590	2	S57594	hypothetical prote	614	7	0.5	687	2	C57713	chloride channel C
542	7	0.5	591	2	S51303	hypothetical prote	615	7	0.5	692	2	S37976	hypothetical prote
543	7	0.5	592	2	B48315	lamin B2 - mouse	616	7	0.5	694	2	T52574	cyclic nucleotide
544	7	0.5	593	2	I51213	drebrin - chicken	617	7	0.5	699	2	A34660	histidine rich cal
545	7	0.5	595	2	A43534	Lupus autoantigen	618	7	0.5	700	2	T05841	spliceosome-assoc1
546	7	0.5	596	2	T23193	hypothetical prote	619	7	0.5	707	1	DNMS	nucleolin - mouse
547	7	0.5	599	2	A45195	adenylcyclase ty	620	7	0.5	708	2	A35804	nucleolin - human
548	7	0.5	596	2	T18316	hypothetical prote	621	7	0.5	707	2	S53411	hypothetical prote
549	7	0.5	599	2	S18735	centromere protein	622	7	0.5	713	2	A27441	nucleolin - Chines
550	7	0.5	602	2	A36715	exo-poly-alpha-gal	623	7	0.5	715	2	D85087	hypothetical prote
551	7	0.5	603	2	T11284	NADH dehydrogenase	624	7	0.5	719	2	S51739	transcription repr
552	7	0.5	603	2	E84744	hypothetical prote	625	7	0.5	720	2	T43327	gluconate transpor
553	7	0.5	606	2	S57552	hypothetical prote	626	7	0.5	721	2	S29795	hypothetical prote
554	7	0.5	606	2	S70358	centromere protein	627	7	0.5	725	2	T42688	hypothetical prote
555	7	0.5	607	2	A43776	drebrin E2 - chick	628	7	0.5	728	2	A54603	transcription fact
556	7	0.5	608	2	T06632	hypothetical prote	629	7	0.5	736	2	T03849	Fas-binding protei
557	7	0.5	610	2	S67701	hypothetical prote	630	7	0.5	740	2	T03847	Fas-binding protei
558	7	0.5	610	2	A96701	protein F12A21.3 [	631	7	0.5	741	2	T40095	zinc finger, C3HC4
559	7	0.5	617	2	S42719	actin-binding prot	632	7	0.5	751	2	A49974	beta-amyloid precu
560	7	0.5	618	1	PRECT4	proteinase IV (EC	633	7	0.5	752	2	S64750	probable Arp-depen
561	7	0.5	618	1	S38004	probable transport	634	7	0.5	757	2	T50959	hypothetical prote
562	7	0.5	618	2	D85786	proteinase IV, a s	635	7	0.5	759	2	S25330	protein T23E23.14
563	7	0.5	618	2	T52076	AB13-interacting p	636	7	0.5	760	2	G86373	amyloid beta (A4)
564	7	0.5	618	2	T52075	pseudo-response re	637	7	0.5	763	2	A49321	probable membrane
565	7	0.5	619	2	E84800	hypothetical prote	638	7	0.5	763	2	S51300	amyloid precursor-
566	7	0.5	620	2	JH0821	95K golgi antigen	639	7	0.5	765	2	S42880	transcription fact
567	7	0.5	622	2	S71342	calnexin precursor	640	7	0.5	765	2	S22314	transcription fact
568	7	0.5	625	2	F70459	conserved hypothet	641	7	0.5	767	2	S41479	DNA-binding protei
569	7	0.5	626	2	T08926	hypothetical prote	642	7	0.5	768	2	H54024	protein kinase (EC
570	7	0.5	627	2	H86180	hypothetical prote	643	7	0.5	771	2	T02565	disease resistance
571	7	0.5	628	2	G71565	probable signal pe	644	7	0.5	772	2	I50463	protein kinase - c
572	7	0.5	629	2	T38214	hypothetical prote	645	7	0.5	776	2	B96666	protein F22C12.6 [
573	7	0.5	629	2	T48799	Rrp9 related prot	646	7	0.5	777	2	B54024	protein kinase (EC
574	7	0.5	630	2	S29796	hypothetical prote	647	7	0.5	777	2	F54024	protein kinase (EC
575	7	0.5	630	2	T48369	hypothetical prote	648	7	0.5	777	2	T08659	ral guanine nucleo
576	7	0.5	631	2	G96701	unknown protein, 7	649	7	0.5	779	2	E54024	protein kinase (EC
577	7	0.5	632	2	T46504	hypothetical prote	650	7	0.5	780	2	F96840	hypothetical prote
578	7	0.5	633	2	T14612	hypothetical prote	651	7	0.5	780	2	I47038	hypothetical prote
579	7	0.5	635	2	H69626	PTS fructose-speci	652	7	0.5	783	2	A55817	cyclin-dependent k
580	7	0.5	637	2	JH0674	L-proline transpor	653	7	0.5	784	2	PN0009	neurofilament trip
581	7	0.5	639	2	T46377	arylsulfatase (EC	654	7	0.5	786	2	A35466	progesterone recep
582	7	0.5	642	2	T23687	hypothetical prote	655	7	0.5	788	2	C70441	lipote-protein 11
583	7	0.5	644	2	S55395	neurofilament prot	656	7	0.5	789	2	T51310	RNA helicase RH28
584	7	0.5	645	1	A23723	protein disulfide-	657	7	0.5	789	2	T52067	hypothetical prote
585	7	0.5	646	2	S55048	protein-tyrosine k	658	7	0.5	792	2	S63141	hypothetical prote
586	7	0.5	646	2	G85056	probable receptor-	659	7	0.5	792	2	T42963	hypothetical prote
587	7	0.5	648	2	F69848	transcription anti	660	7	0.5	794	2	T27870	hypothetical prote
588	7	0.5	649	2	T39826	probable clathrin	661	7	0.5	798	2	D96563	probable bZIP prot
589	7	0.5	651	2	T15624	hypothetical prote	662	7	0.5	798	2	T33022	hypothetical prote
590	7	0.5	653	2	A39922	potassium channel	663	7	0.5	798	2	JC7500	glik protein - chic
591	7	0.5	654	2	S11049	potassium channel	664	7	0.5	799	2	JH0797	castor protein - f
592	7	0.5	660	2	B96555	hypothetical prote	665	7	0.5	802	1	S48529	NAB3 protein - yea
593	7	0.5	661	2	T24584	hypothetical prote	666	7	0.5	806	2	S22765	heterogeneous ribo
594	7	0.5	661	2	A81125	bacteriophage tran	667	7	0.5	808	2	T23129	hypothetical prote
595	7	0.5	664	2	T05584	probable serine-th	668	7	0.5	809	1	JQ0032	anthrax toxin leth
596	7	0.5	668	2	C70527	probable disease r	669	7	0.5	809	2	S67665	ubiquitin-specific
597	7	0.5	669	2	S78050	high mobility grou	670	7	0.5	817	2	JC4176	pyruvate,water dik
598	7	0.5	672	2	T32557	hypothetical prote	671	7	0.5	819	1	S40400	protein kinase SWE
599	7	0.5	674	2	F81990	NADH dehydrogenase	672	7	0.5	819	2	F75196	pyruvate,water dik
600	7	0.5	674	2	D81220	NADH dehydrogenase	673	7	0.5	820	2	T00645	hypothetical prote
601	7	0.5	674	2	D82118	conserved hypothet	674	7	0.5	821	2	B71229	pyruvate,water dik
602	7	0.5	675	2	T03744	myoD protein inhib	675	7	0.5	825	2	S54465	YTA12 protein prec
603	7	0.5	676	2	T01084	hypothetical prote	676	7	0.5	830	2	T07743	probable 1,4-alpha
604	7	0.5	678	2	A54514	glutamic acid-rich	677	7	0.5	845	2	JC5256	adipocyte transcri
605	7	0.5	678	2	T43539	spindle checkpoint	678	7	0.5	852	2	A34373	histidine-rich cal
606	7	0.5	679	2	S48939	hypothetical prote	679	7	0.5	855	2	T10665	hypothetical prote
607	7	0.5	679	2	A83624	probable chemotaxi	680	7	0.5	858	2	T08881	prominin - mouse
608	7	0.5	680	2	T19768	hypothetical prote	681	7	0.5	859	2	C96504	protein F9C16.23 [
609	7	0.5	683	2	D71433	hypothetical prote	682	7	0.5	862	2	T34342	hypothetical prote
610	7	0.5	684	2	T51410	N2,N2-dimethylguan	683	7	0.5	883	2	A71434	probable RNA helic
611	7	0.5	686	2	A45483	chloride channel,	684	7	0.5	884	2	S66308	nitrate reductase
612	7	0.5	687	2	A57713	chloride channel C	685	7	0.5	886	1	RGBYL3	regulatory protein
613	7	0.5	687	2	B57713	chloride channel K	686	7	0.5	886	2	A59223	nitrate reductase



395	7	0.5	427	2	S38002	hypothetical prote	468	7	0.5	495	2	S10851	glycinin G1 precu
396	7	0.5	429	2	I49603	transcription regu	469	7	0.5	495	2	C71410	hypothetical prote
397	7	0.5	429	2	F86240	hypothetical prote	470	7	0.5	496	2	S68160	probable RNA bindi
398	7	0.5	430	2	I57013	guanine nucleotide	471	7	0.5	497	2	JC5076	myc-associated zin
399	7	0.5	431	2	C69087	dihydrolipoamide d	472	7	0.5	498	2	T48385	transporter like p
400	7	0.5	433	2	JC4529	guanylate cyclase	473	7	0.5	500	2	S26688	legumin K - garden
401	7	0.5	433	2	D53352	deoxyguanosinetrip	474	7	0.5	500	2	T42901	hypothetical prote
402	7	0.5	434	2	C71418	cytochrome P450 -	475	7	0.5	502	2	I52637	Ca2+/calmodulin-de
403	7	0.5	435	2	S31290	cyclin B5 - yeast	476	7	0.5	502	2	S01401	H+-transporting AT
404	7	0.5	435	2	T01826	microfibril-associ	477	7	0.5	502	2	F31482	H+-transporting AT
405	7	0.5	436	2	T38812	hypothetical prote	478	7	0.5	502	2	T48560	hypothetical prote
406	7	0.5	440	2	S17195	transcription fact	479	7	0.5	502	2	T01179	hypothetical prote
407	7	0.5	445	2	S43328	tubulin beta-7 cha	480	7	0.5	503	2	S00336	legumin B legu pre
408	7	0.5	445	2	T47813	hypothetical prote	481	7	0.5	504	1	A32965	cytochrome P450 4A
409	7	0.5	446	2	A32284	chromogranin A pre	482	7	0.5	504	2	I53868	alpha-internexin -
410	7	0.5	448	2	T06698	hypothetical prote	483	7	0.5	504	2	G02474	interferon regulat
411	7	0.5	448	2	S05355	hypothetical prote	484	7	0.5	504	2	T10558	hypothetical prote
412	7	0.5	449	2	J8A049	tubulin beta-2 cha	485	7	0.5	505	2	A41023	alpha-internexin -
413	7	0.5	449	2	A5709	major early-transc	486	7	0.5	507	1	A32966	cytochrome P450 4A
414	7	0.5	450	2	A37223	alpha-2B-adrenergi	487	7	0.5	509	2	S18872	legumin-like stora
415	7	0.5	450	2	T10632	hypothetical prote	488	7	0.5	509	2	E86427	hypothetical prote
416	7	0.5	452	2	T20556	hypothetical prote	489	7	0.5	511	2	A40110	nicotinic acetylch
417	7	0.5	452	2	I39383	angio-associated m	490	7	0.5	511	2	JC1404	CDE1-box DNA-bindi
418	7	0.5	454	2	H86231	hypothetical prote	491	7	0.5	516	2	T48394	hypothetical prote
419	7	0.5	455	1	B25493	indoleacetamide hy	492	7	0.5	517	2	T23499	hypothetical prote
420	7	0.5	455	2	S65157	hypothetical prote	493	7	0.5	517	2	A49413	perilipin A - rat
421	7	0.5	455	2	I48724	zinc finger protei	494	7	0.5	518	2	T24783	hypothetical prote
422	7	0.5	456	2	D84651	hypothetical prote	495	7	0.5	522	2	JC4024	poliovirus recepto
423	7	0.5	457	1	A28468	chromogranin A pre	496	7	0.5	522	2	A55194	importin 1 - Afric
424	7	0.5	461	2	B93601	probable transport	497	7	0.5	523	2	B55194	importin 2 - Afric
425	7	0.5	462	2	A40552	bindin fertilizati	498	7	0.5	523	2	S67271	hypothetical prote
426	7	0.5	463	2	S28469	phosphomannomutase	499	7	0.5	524	2	A86356	hypothetical prote
427	7	0.5	464	2	S45363	LEO1 protein - yea	500	7	0.5	526	2	G83436	hypothetical prote
428	7	0.5	466	2	T22141	hypothetical prote	501	7	0.5	527	1	A71418	cytochrome P450 d1
429	7	0.5	467	2	G84937	glutamate--trNA li	502	7	0.5	527	2	A32469	80K protein H prec
430	7	0.5	467	2	S61105	hypothetical prote	503	7	0.5	528	1	ACCH2N	nicotinic acetylch
431	7	0.5	470	2	S47887	BF11 protein - yea	504	7	0.5	531	2	A55887	caldesmon, non-mus
432	7	0.5	471	2	B83127	probable secretion	505	7	0.5	531	2	T33980	hypothetical prote
433	7	0.5	471	2	T49019	probable RNA bindi	506	7	0.5	531	2	G83268	conserved hypothet
434	7	0.5	473	2	S22321	deoxyribodipyrimid	507	7	0.5	531	2	T49058	hypothetical prote
435	7	0.5	474	1	TVR7C4	Ca2+/calmodulin-de	508	7	0.5	533	2	T34458	hypothetical prote
436	7	0.5	474	2	S28419	lamin B-3 - mouse	509	7	0.5	534	2	S61185	hypothetical prote
437	7	0.5	475	2	A71110	hypothetical prote	510	7	0.5	538	2	T03262	cytochrome P450 -
438	7	0.5	475	2	S28933	hypothetical prote	511	7	0.5	541	2	T48811	hypothetical prote
439	7	0.5	475	2	C86863	N-acetylmuramoyl-L	512	7	0.5	544	2	A57597	beta-p21-activated
440	7	0.5	477	2	F85169	cytochrome P450 li	513	7	0.5	544	2	I49376	p21 activated kina
441	7	0.5	477	2	A47236	zinc-finger protei	514	7	0.5	545	2	H86322	calcium-dependent
442	7	0.5	478	2	I53960	PRR2 alpha - human	515	7	0.5	545	2	S68687	thermosome - Metha
443	7	0.5	478	2	S56904	hypothetical prote	516	7	0.5	547	2	T45635	hypothetical prote
444	7	0.5	480	4	TVHUBD	transforming prote	517	7	0.5	550	2	B49000	hypothetical prote
445	7	0.5	480	2	S14394	bindin precursor -	518	7	0.5	552	2	S45886	hypothetical prote
446	7	0.5	481	2	T18465	hypothetical prote	519	7	0.5	552	2	H70445	hypothetical prote
447	7	0.5	481	2	A26483	bindin precursor -	520	7	0.5	553	2	T06499	hypothetical prote
448	7	0.5	482	1	XUBOLA	dihydrolipoamide S	521	7	0.5	554	2	A56730	carl protein - Pod
449	7	0.5	482	2	A32422	dihydrolipoamide S	522	7	0.5	559	1	WMBES1	71k upper matrix p
450	7	0.5	482	2	S65760	dihydrolipoamide t	523	7	0.5	559	2	T48180	hypothetical prote
451	7	0.5	482	2	S37845	transcription init	524	7	0.5	560	2	T02404	probable beta-gluc
452	7	0.5	483	2	S27880	Nasopressin recept	525	7	0.5	561	2	JC2436	5'-nucleotidase (E
453	7	0.5	484	2	A24942	legumin B4 precurs	526	7	0.5	566	2	T06453	probable legumin B
454	7	0.5	484	2	T06063	hypothetical prote	527	7	0.5	566	2	S28068	mx1D protein - Shi
455	7	0.5	484	2	S61870	hypothetical prote	528	7	0.5	569	2	S56658	stress-induced pro
456	7	0.5	485	2	B40552	bindin fertilizati	529	7	0.5	570	1	B38546	major structural n
457	7	0.5	486	2	JC7241	calcineurin cataly	530	7	0.5	572	2	T22547	hypothetical prote
458	7	0.5	486	2	S96649	hypothetical prote	531	7	0.5	580	2	T43485	hypothetical prote
459	7	0.5	486	2	S67589	probable membrane	532	7	0.5	580	2	G86169	hypothetical prote
460	7	0.5	491	2	H84477	probable PttA-like	533	7	0.5	581	2	T46024	hypothetical prote
461	7	0.5	491	2	S49790	probable membrane	534	7	0.5	581	2	C96538	hypothetical prote
462	7	0.5	492	2	A61382	phosphorylation re	535	7	0.5	582	2	S53814	DEAD box protein -
463	7	0.5	493	1	S46283	calcium-dependent	536	7	0.5	582	2	S19424	hypothetical prote
464	7	0.5	493	2	A26579	cytochrome P450 2E	537	7	0.5	584	2	T23174	hypothetical prote
465	7	0.5	493	2	G84594	probable diacylgly	538	7	0.5	586	2	C30411	synapsin Iia - rat
466	7	0.5	494	2	A42170	zinc finger protei	539	7	0.5	587	2	B83015	arginyl-tRNA synth
467	7	0.5	495	1	FWSYG2	glycinin chain Ala	540	7	0.5	587	2	JC5300	Ran GTPase activat



249	7	0.5	280	2	S35103	bone sialoprotein	322	7	0.5	340	2	G82047	conserved hypothet
250	7	0.5	280	2	H72389	transcription regu	323	7	0.5	340	2	T51386	probable protein w
251	7	0.5	282	1	D64757	ykG protein - Esc	324	7	0.5	341	2	T41450	hypothetical prote
252	7	0.5	284	2	B48516	surfactant protein	325	7	0.5	343	2	S45321	folliculin - mous
253	7	0.5	288	2	I51332	signal sequence re	326	7	0.5	343	2	T02399	hypothetical prote
254	7	0.5	288	2	T32459	hypothetical prote	327	7	0.5	344	2	I57698	folliculin - rat
255	7	0.5	289	2	JC7279	Down syndrome crit	328	7	0.5	347	2	T33204	hypothetical prote
256	7	0.5	292	2	E72223	conserved hypothet	329	7	0.5	347	2	S40105	gene AGP2beta-2 pr
257	7	0.5	292	2	A84427	hypothetical prote	330	7	0.5	350	2	S00337	legumin B Legk pre
258	7	0.5	292	2	T00996	En/Spm-like transp	331	7	0.5	350	2	T06107	probable serine/th
259	7	0.5	293	2	T09170	ribosomal protein	332	7	0.5	350	2	T33458	hypothetical prote
260	7	0.5	295	2	A38977	cyclin D1 - human	333	7	0.5	351	2	F96597	hypothetical prote
261	7	0.5	295	2	T39012	hypothetical prote	334	7	0.5	353	2	A86438	hypothetical prote
262	7	0.5	295	2	E84862	hypothetical prote	335	7	0.5	354	2	I48722	zinc finger protei
263	7	0.5	296	2	G86523	hypothetical prote	336	7	0.5	357	2	JC4703	basic helix-loop-h
264	7	0.5	296	2	G72099	conserved hypothet	337	7	0.5	357	2	I49338	neurogenic differe
265	7	0.5	298	2	A71159	probable membrane	338	7	0.5	357	2	D71262	probable hemolysin
266	7	0.5	298	2	C86871	conserved hypothet	339	7	0.5	359	2	T21840	hypothetical prote
267	7	0.5	299	2	F75003	stomatol-like prot	340	7	0.5	363	2	D64640	hypothetical prote
268	7	0.5	299	2	A54090	pXf protein - Chin	341	7	0.5	366	2	A96692	hypothetical prote
269	7	0.5	299	2	I37468	house keeping gene	342	7	0.5	368	1	TVM5ML	transforming prote
270	7	0.5	299	2	H84661	heme oxygenase 2 (	343	7	0.5	368	2	H96712	probable DNA-bindi
271	7	0.5	299	2	T02515	cytoskeletal prote	344	7	0.5	369	2	D81984	riboflavin bifunct
272	7	0.5	301	2	T20651	hypothetical prote	345	7	0.5	371	2	T21707	hypothetical prote
273	7	0.5	301	2	JW0079	heterogeneous nucl	346	7	0.5	372	2	T02550	NPk1-related prote
274	7	0.5	302	2	E86267	hypothetical prote	347	7	0.5	372	2	S76427	hypothetical prote
275	7	0.5	303	1	S28392	protein-tyrosine p	348	7	0.5	375	2	A64348	hypothetical prote
276	7	0.5	303	2	S12867	carbonate dehydrat	349	7	0.5	375	2	S26059	probable transform
277	7	0.5	303	2	T10653	hypothetical prote	350	7	0.5	376	2	T34734	hypothetical prote
278	7	0.5	303	2	T48493	hypothetical prote	351	7	0.5	376	2	T10455	heat shock related
279	7	0.5	304	2	S44897	ZK1236.2 protein -	352	7	0.5	381	2	G02668	neurogenic basic-h
280	7	0.5	306	2	G70481	thiamin monophosph	353	7	0.5	381	2	JC4647	KW8 protein - rat
281	7	0.5	306	2	T32461	hypothetical prote	354	7	0.5	381	2	T49544	hypothetical prote
282	7	0.5	310	2	I46987	bone sialoprotein	355	7	0.5	382	2	T11219	ubiquinol--cytochr
283	7	0.5	312	2	S19378	hypothetical prote	356	7	0.5	382	2	S47747	2-dehydro-3-deoxy
284	7	0.5	312	2	S66952	hypothetical prote	357	7	0.5	382	2	S72169	DNA-binding protei
285	7	0.5	314	2	T16300	hypothetical prote	358	7	0.5	382	2	G86025	ketodeoxygluconoki
286	7	0.5	316	2	T11021	farnesyltransferra	359	7	0.5	383	1	A48222	dematin 48k chain
287	7	0.5	316	2	G96513	hypothetical prote	360	7	0.5	383	2	JC4688	neuro D-related fa
288	7	0.5	316	2	S73683	MG338 homolog P02-	361	7	0.5	386	2	S41497	thyroid hormone re
289	7	0.5	317	1	GEH091	bone sialoprotein	362	7	0.5	387	2	A83393	probable acyl-CoA
290	7	0.5	317	2	S67493	huntingtin-associa	363	7	0.5	387	2	S02708	troponin T - fruit
291	7	0.5	320	2	T49591	probable 35 kDa ri	364	7	0.5	388	3	T00641	hypothetical prote
292	7	0.5	320	1	GERTS	bone sialoprotein	365	7	0.5	388	3	JC7510	benzoate X recepto
293	7	0.5	320	2	S72555	huntingtin-associa	366	7	0.5	392	1	S55971	probable peptidylp
294	7	0.5	323	2	F70646	probable quinone o	367	7	0.5	393	2	T43401	transcription init
295	7	0.5	324	2	I49768	bone sialoprotein	368	7	0.5	394	2	S64505	HGH1 protein - yea
296	7	0.5	324	2	D72070	conserved hypothet	369	7	0.5	396	2	F75006	hypothetical prote
297	7	0.5	324	2	B86555	CT429 hypothetical	370	7	0.5	396	2	S13251	troponin T - fruit
298	7	0.5	327	2	G33282	DNA-binding protei	371	7	0.5	400	2	A26258	endoplasmin - gold
299	7	0.5	328	2	S71425	DNA-directed RNA p	372	7	0.5	400	2	A55647	phyllorod - fruit
300	7	0.5	328	2	B84545	hypothetical prote	373	7	0.5	405	1	JH0795	calreticulin precu
301	7	0.5	329	1	RNECA	DNA-directed RNA p	374	7	0.5	405	2	I39062	dematin 52K chain
302	7	0.5	329	1	A41658	DNA-directed RNA p	375	7	0.5	406	2	A48059	oncoprotein zc-Myc
303	7	0.5	329	1	E72068	carbohydrate isome	376	7	0.5	406	2	S59948	aminomethyltransf
304	7	0.5	329	2	H84987	DNA-directed RNA p	377	7	0.5	409	2	E86336	hypothetical prote
305	7	0.5	329	2	D85994	RNA polymerase, al	378	7	0.5	411	2	S48647	peptidylprolyl iso
306	7	0.5	329	2	S07577	legumin storage pr	379	7	0.5	412	2	A55320	immunophilin FRB4
307	7	0.5	329	2	B86556	GutQ/KpsF family p	380	7	0.5	412	2	B86391	hypothetical prote
308	7	0.5	329	2	E71516	hypothetical prote	381	7	0.5	412	2	G02453	NN8-4AG - human (f
309	7	0.5	331	2	T41507	hypothetical prote	382	7	0.5	414	2	T49459	hypothetical prote
310	7	0.5	331	2	B47236	zinc-finger protei	383	7	0.5	415	2	JC7167	C kinase 1 interac
311	7	0.5	332	2	A81675	conserved hypothet	384	7	0.5	416	1	A42879	advanced glycosyla
312	7	0.5	332	2	E86448	hypothetical prote	385	7	0.5	416	2	T10623	hypothetical prote
313	7	0.5	333	2	D83113	DNA-directed RNA p	386	7	0.5	417	2	H72362	3-isopropylmalate
314	7	0.5	335	2	S07576	legumin storage pr	387	7	0.5	417	2	A49129	neuroblast cell li
315	7	0.5	335	2	B86429	hypothetical prote	388	7	0.5	420	2	B86217	protein T27G7.4 [i
316	7	0.5	336	2	G69091	ribosomal protein	389	7	0.5	420	2	T46910	hypothetical prote
317	7	0.5	336	2	F71942	holliday junction	390	7	0.5	422	2	T41888	hypothetical prote
318	7	0.5	336	2	C64652	DNA-binding protei	391	7	0.5	423	2	B86214	IE-2 (IE-N) orf151
319	7	0.5	337	2	S31131	hypothetical prote	392	7	0.5	424	2	T01383	GTPase-activating
320	7	0.5	338	2	S04321	legumin B (clone p	393	7	0.5	424	2	S43560	coiled coil protei
321	7	0.5	339	2	B86281	protein F10B6.6 [i	394	7	0.5	426	2	I48379	gene hb protein -



103	7	0.5	116	2	PC4228	N-acetylglucosamin	176	7	0.5	210	2	A34719	nonhistone chromos
104	7	0.5	118	2	T05337	hypothetical protei	177	7	0.5	210	2	S54774	high mobility grou
105	7	0.5	119	2	A03314	homeotic protein m	178	7	0.5	211	2	I53020	G-0/G-1 switch reg
106	7	0.5	120	2	T08717	hypothetical prote	179	7	0.5	211	2	S69929	outer surface prot
107	7	0.5	121	2	A70200	ribosome-binding f	180	7	0.5	213	2	J00075	neuromodulin - gol
108	7	0.5	129	2	H84749	probable squamosa-	181	7	0.5	213	2	J00931	DNA binding protei
109	7	0.5	130	2	G70973	hypothetical prote	182	7	0.5	213	2	G86261	hypothetical prote
110	7	0.5	131	2	P00059	T-cell receptor be	183	7	0.5	214	2	G69380	hypothetical prote
111	7	0.5	131	2	E70920	probable moab prot	184	7	0.5	214	2	T51657	myb-related trans
112	7	0.5	131	2	T52597	squamosa promoter	185	7	0.5	214	2	T39559	probable ubiquinol
113	7	0.5	133	2	H59489	SSU ribosomal prot	186	7	0.5	215	1	NSRTH1	nonhistone chromos
114	7	0.5	134	2	E57233	complexin II - hum	187	7	0.5	215	2	I48688	non-histone chromo
115	7	0.5	134	2	D57233	complexin II - mou	188	7	0.5	215	2	T05158	hypothetical prote
116	7	0.5	134	2	JC4226	synaphin - bovine	189	7	0.5	215	2	S29595	calmodulin-related
117	7	0.5	134	2	C57233	complexin II - rat	190	7	0.5	216	2	D84841	calmodulin-like pr
118	7	0.5	134	2	T75416	hypothetical prote	191	7	0.5	217	2	T48642	hypothetical prote
119	7	0.5	136	2	T30045	hypothetical prote	192	7	0.5	219	2	C96510	hypothetical prote
120	7	0.5	138	2	T16473	hypothetical prote	193	7	0.5	220	2	S01063	Hox 2 protein type
121	7	0.5	140	2	G84608	En/spm-like transp	194	7	0.5	221	2	A86163	protein F15K9.22 l
122	7	0.5	141	2	T46015	hypothetical prote	195	7	0.5	223	2	T46043	hypothetical prote
123	7	0.5	142	2	S63373	probable membrane	196	7	0.5	224	2	S51093	transcription acti
124	7	0.5	143	2	T29740	hypothetical prote	197	7	0.5	224	2	JH0456	tumor rejection an
125	7	0.5	144	2	B72280	hypothetical prote	198	7	0.5	224	2	S68150	eosinophil major b
126	7	0.5	149	2	F83964	hypothetical prote	199	7	0.5	228	2	C89347	hypothetical prote
127	7	0.5	152	2	S23311	HMG-Y-related prot	200	7	0.5	229	2	A28329	homeotic protein H
128	7	0.5	154	2	T28434	transcription regu	201	7	0.5	230	2	T39119	protein-L-isoaspar
129	7	0.5	155	2	T05169	Lsd1 protein homol	202	7	0.5	232	2	A69501	hypothetical prote
130	7	0.5	157	2	F71906	hypothetical prote	203	7	0.5	233	2	H82995	hypothetical prote
131	7	0.5	159	2	T44144	B1 protein (import	204	7	0.5	234	1	S15102	conserved hypotet
132	7	0.5	161	2	G89173	hypothetical prote	205	7	0.5	234	2	B44459	eosinophil major b
133	7	0.5	170	1	WNADE5	early E1B 20K prot	206	7	0.5	235	2	F81322	tropomycin T, fast s
134	7	0.5	170	2	A27853	nonhistone chromos	207	7	0.5	235	2	S33662	hypothetical prote
135	7	0.5	170	2	S69465	hypothetical prote	208	7	0.5	236	2	S33662	hypothetical prote
136	7	0.5	172	2	D72368	hypothetical prote	209	7	0.5	238	1	Q0BE74	US4 protein - huma
137	7	0.5	173	2	JC6004	superoxide dismuta	210	7	0.5	241	2	A84517	Mutator-like trans
138	7	0.5	173	2	G85771	superoxide dismuta	211	7	0.5	241	2	T00988	hypothetical prote
139	7	0.5	173	2	S73388	hypothetical prote	212	7	0.5	242	2	S19267	anthranilate phosp
140	7	0.5	173	2	T08011	2S seed storage pr	213	7	0.5	246	2	S38344	CDEI-binding prote
141	7	0.5	176	2	S23310	HMG-Y-related prot	214	7	0.5	247	2	T05226	hypothetical prote
142	7	0.5	177	2	T49816	hypothetical prote	215	7	0.5	248	2	T02554	hypothetical prote
143	7	0.5	177	2	C72765	hypothetical prote	216	7	0.5	248	2	T05080	hypothetical prote
144	7	0.5	178	2	T09585	high mobility grou	217	7	0.5	249	2	T44220	hypothetical prote
145	7	0.5	178	2	T09584	high mobility grou	218	7	0.5	249	2	A84751	probable SF16 prot
146	7	0.5	178	2	T51159	HMG protein (import	219	7	0.5	250	2	C71677	hypothetical prote
147	7	0.5	178	2	C83705	hypothetical prote	220	7	0.5	250	2	T26010	hypothetical prote
148	7	0.5	181	1	ERAD23	early E3 20.3K gly	221	7	0.5	252	2	T01109	hypothetical prote
149	7	0.5	181	1	B44057	early E3 20.3K gly	222	7	0.5	256	2	A70637	hypothetical prote
150	7	0.5	181	1	F44057	early E3 20.3K gly	223	7	0.5	256	2	S37926	hypothetical prote
151	7	0.5	181	2	T15117	hypothetical prote	224	7	0.5	259	2	G83675	RNA-binding protei
152	7	0.5	183	2	A84122	pyrazinamidase/nic	225	7	0.5	259	2	I50727	sensory organ home
153	7	0.5	184	2	A34506	23K ubiquitin carr	226	7	0.5	260	2	T25017	hypothetical prote
154	7	0.5	184	2	D84746	hypothetical prote	227	7	0.5	261	2	S10321	31K protein - frog
155	7	0.5	186	2	S30221	nonhistone chromos	228	7	0.5	261	2	B86358	hypothetical prote
156	7	0.5	190	2	A26630	nucleoplasmin A -	229	7	0.5	262	2	S42540	ribosomal protein
157	7	0.5	191	2	G71710	sco2 protein precu	230	7	0.5	262	2	F83498	hypothetical prote
158	7	0.5	193	2	S70260	outer surface prot	231	7	0.5	262	2	T41536	yeast anti-silenci
159	7	0.5	193	2	S70266	outer surface prot	232	7	0.5	264	2	B96680	hypothetical prote
160	7	0.5	194	2	H83060	peptidyl-L-TRNA hydr	233	7	0.5	265	2	H83857	tryptophan synthas
161	7	0.5	196	2	D82143	conserved hypotet	234	7	0.5	265	2	JC4970	slow muscle tropon
162	7	0.5	197	2	S65050	low molecular weig	235	7	0.5	265	2	T46013	hypothetical prote
163	7	0.5	200	2	A26169	nucleoplasmin - Af	236	7	0.5	265	2	A43738	Endo16 protein - s
164	7	0.5	200	2	G71405	hypothetical prote	237	7	0.5	266	2	H86407	hypothetical prote
165	7	0.5	202	2	A26301	dopamine- and camp	238	7	0.5	267	1	JQ1323	DNA excision repai
166	7	0.5	203	2	T50421	nonhistone chromos	239	7	0.5	267	2	H86320	probable MYB47 tra
167	7	0.5	205	2	S26062	nonhistone chromos	240	7	0.5	267	2	H84857	hypothetical prote
168	7	0.5	205	2	H59889	hypothetical prote	241	7	0.5	267	2	E83232	probable ATP-bindi
169	7	0.5	206	2	S30108	S-6b protein - Agr	242	7	0.5	269	2	T16487	hypothetical prote
170	7	0.5	206	2	A69712	mutants block spor	243	7	0.5	271	2	G85035	hypothetical prote
171	7	0.5	208	2	S46510	hypothetical prote	244	7	0.5	271	2	D82314	transcription regu
172	7	0.5	208	2	T05871	hypothetical prote	245	7	0.5	273	1	I38886	DNA excision repai
173	7	0.5	209	1	NSHUH2	nonhistone chromos	246	7	0.5	273	2	JG0190	XPA protein - huma
174	7	0.5	209	2	C69783	NADH dehydrogenase	247	7	0.5	276	2	I53161	alpha 2-adrenergic
175	7	0.5	209	2	S89275	CGMP-gated cation	248	7	0.5	278	2	S48776	hypothetical prote
										279	2	S30766	ASF1 protein - yea



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:04:14 ; Search time 26.41 Seconds  
(without alignments)  
4012.071 Million cell updates/sec

Title: US-09-512-581-2

Perfect score: 1391

Sequence: 1 MAHSKTRNDGKITYPEGVK.....OKGRGRPSKTPSPQPKKNV 1391

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	19	1.4	851	2	T00374	hypothetical prote
2	9	0.6	439	2	A57573	telomeric repeat b
3	9	0.6	554	2	T45840	GTPase activating-
4	9	0.6	837	2	T49562	neurofilament-H re
5	9	0.6	1938	1	S06005	myosin alpha heavy
6	9	0.6	1938	2	I49464	alpha cardiac myos
7	8	0.6	81	2	T47289	hypothetical prote
8	8	0.6	87	2	S27275	GTP-binding regula
9	8	0.6	91	1	S00219	ubiquinol--cytochr
10	8	0.6	99	2	A55819	nonhistone chromos
11	8	0.6	147	1	B46315	E4 protein - huma
12	8	0.6	204	2	T05575	hypothetical prote
13	8	0.6	207	2	JC1129	nonhistone chromos
14	8	0.6	231	2	JC1114	high-mobility grou
15	8	0.6	231	2	T49830	hypothetical prote
16	8	0.6	233	1	S13625	eosinophil major b
17	8	0.6	233	2	A23729	somatolactin precu
18	8	0.6	242	1	RHHUT	thyroliberin precu
19	8	0.6	294	2	E84706	hypothetical prote
20	8	0.6	313	2	F84923	pale cress protein
21	8	0.6	316	2	S58719	probable membrane
22	8	0.6	316	2	JC6146	Card protein - Myx
23	8	0.6	325	2	T18283	hypothetical prote
24	8	0.6	339	2	B69436	LSU ribosomal prot
25	8	0.6	341	2	G72775	hypothetical prote
26	8	0.6	351	2	B71604	rRNA methylase (Sp
27	8	0.6	398	2	S06324	dnaB protein homol
28	8	0.6	418	2	T15827	hypothetical prote
29	8	0.6	437	2	H86683	prophage pil prote

30	8	0.6	442	1	E70046	iron transport sys
31	8	0.6	451	2	A37386	dnaB protein homol
32	8	0.6	451	2	S01921	dnaB protein homol
33	8	0.6	462	1	A60746	chromogranin A pre
34	8	0.6	463	1	A39868	chromogranin A pre
35	8	0.6	468	2	G71231	probable replicati
36	8	0.6	507	2	F84558	hypothetical prote
37	8	0.6	513	2	T10830	nitrogenase (EC 1.
38	8	0.6	513	2	T00828	hypothetical prote
39	8	0.6	533	2	A56110	tyrosine phosphopr
40	8	0.6	544	1	I36911	involucrin L - dou
41	8	0.6	544	2	B44841	low molecular weig
42	8	0.6	552	2	T04653	receiver-like prot
43	8	0.6	562	2	T49788	related to merzoi
44	8	0.6	564	2	S37241	legumin B - fava b
45	8	0.6	566	2	A72329	general secretion
46	8	0.6	568	2	H86552	prolyl tRNA synthe
47	8	0.6	568	2	F72070	proline--tRNA liga
48	8	0.6	575	2	A96766	unknown protein F2
49	8	0.6	586	1	VEHULB	lamin B1 - human
50	8	0.6	587	2	S07720	lamin B - mouse
51	8	0.6	646	2	T02398	hypothetical prote
52	8	0.6	653	2	A46362	amyloid precursor-
53	8	0.6	655	2	F82138	GGDEF family prote
54	8	0.6	671	2	S61693	probable membrane
55	8	0.6	720	2	T26819	hypothetical prote
56	8	0.6	730	2	C84523	Ac-like transposas
57	8	0.6	765	2	E96558	hypothetical prote
58	8	0.6	787	2	S68699	potassium channel
59	8	0.6	841	2	B71212	hypothetical prote
60	8	0.6	938	2	B57487	inositol-polyphosp
61	8	0.6	1002	2	S70292	FUN12 protein - ye
62	8	0.6	1047	2	T49425	hypothetical prote
63	8	0.6	1062	2	T46444	hypothetical prote
64	8	0.6	1081	2	T09837	sucrose-phosphate
65	8	0.6	1082	2	T42204	chromatin structur
66	8	0.6	1132	2	T43483	translation initia
67	8	0.6	1165	2	A6180	adenyl cyclase t
68	8	0.6	1166	2	A49201	adenyl cyclase
69	8	0.6	1173	2	T31421	C-terminal domain-
70	8	0.6	1180	2	A47202	adenylate cyclase
71	8	0.6	1184	2	A55184	fibulin-2 precurs
72	8	0.6	1231	2	T24415	hypothetical prote
73	8	0.6	1298	2	I54367	X-linked nuclear p
74	8	0.6	1634	2	T26517	hypothetical prote
75	8	0.6	1841	2	I38614	helicase II - huma
76	8	0.6	1736	2	T00391	hypothetical prote
77	8	0.6	1871	2	D96796	probable heat shoc
78	8	0.6	1934	2	I48153	myosin heavy chain
79	8	0.6	1935	1	S06006	myosin beta heavy
80	8	0.6	2248	1	D42088	adenylate cyclase
81	8	0.6	4868	2	B54161	ryanodine-binding
82	7	0.5	35	2	T00418	hypothetical prote
83	7	0.5	41	2	E70210	hypothetical prote
84	7	0.5	46	2	S25149	Ku antigen, 70K -
85	7	0.5	75	2	T06013	hypothetical prote
86	7	0.5	78	1	CCB011	ubiquinol--cytochr
87	7	0.5	81	2	T48398	hypothetical prote
88	7	0.5	91	2	T45132	hypothetical prote
89	7	0.5	95	2	T46086	hypothetical prote
90	7	0.5	101	2	H69091	ribosomal protein
91	7	0.5	102	2	B31512	parathymosin - rat
92	7	0.5	105	2	A27471	homeotic protein R
93	7	0.5	106	2	S20422	zinc-binding prote
94	7	0.5	106	2	C69436	LSU ribosomal prot
95	7	0.5	107	2	C86477	protein F1504.29 l
96	7	0.5	109	1	TNBOA1	prothymosin alpha
97	7	0.5	109	2	F72666	prothymosin alpha
98	7	0.5	110	1	TNBUA	prothymosin alpha
99	7	0.5	111	2	S15073	prothymosin alpha
100	7	0.5	112	1	TNRTA	prothymosin alpha
101	7	0.5	112	2	T18436	hypothetical prote
102	7	0.5	112	2	S57448	DNA binding protei







ACCESSION AL358892  
VERSION AL358892.12 GI:11493289  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 195380)  
Wall, M.  
REFERENCE Direct Submission  
AUTHORS Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
JOURNAL CE10 ISA, UK. E-mail enquiries: humquereyesanger.ac.uk  
On Nov 30, 2000 this sequence version replaced gi:11414592.  
----- Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: http://mrcseq.har.mrc.ac.uk  
Contact: mouse@har.mrc.ac.uk  
----- Project Information  
Center project name: dm417G6  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 190612 bases at least Q40  
Consensus quality: 192224 bases at least Q30  
Consensus quality: 193166 bases at least Q20  
Insert size: 194680; sum-of-contigs  
Insert size: 191947; 3.7% error; agarose-fp  
Quality coverage: 7.33x in Q20 bases; sum-of-contigs quality  
coverage: 8.85x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----  
FEATURES  
source location/qualifiers  
1. 195380  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/clone="RP21-417G6"  
/clone\_1id="RPCI-21"  
1. 2333  
/note="assembly\_fragment:02826  
fragment\_chain:1"  
misc\_feature 2434..4472  
/note="assembly\_fragment:00159  
fragment\_chain:1"  
misc\_feature 4573..6623  
/note="assembly\_fragment:00292"  
6724..8757  
/note="assembly\_fragment:02097"  
8858..11498  
/note="assembly\_fragment:02560"  
11599..13673  
/note="assembly\_fragment:03440"  
13774..16148  
/note="assembly\_fragment:03977"  
16249..195380  
/note="assembly\_fragment:04183"  
BASE COUNT 55503 a 40658 c 40664 g 58048 t 707 others  
ORIGIN

Query Match 4.3%; Score 179.6; DB 80; Length 195380;  
Best Local Similarity 91.0%; Pred. No. 8.3e-27;  
Matches 191; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 104 taagaatggtgtgaaacctttatgatgatgacacgaagactctgaagaagaagaagc 163  
|||||  
Db 168033 TATAGATGTGTAAACCTTCAATGACATGACAGACTCTGAGAGAAAGAAAC 168092

QY 164 ttattaaacctagcttatacatctgtcagattttttcacaagacatcgtgtaaag 223  
|||||  
Db 168093 TTTATCTAACCCTAGCTTTACATCTTGCTTGACTTCTCCTCAAGCATCTGTATAAG 168152  
|||||

QY 224 atgttcgttactgtgtagcctgcgccttctgatatattcagagattatgctcctgaag 283  
|||||  
Db 168153 ATGTCGTTTACTGTGTGCTTCTCCTGCTGATATTTCAGCATTTATCTCTCTGAGG 168212  
|||||

QY 284 ctcccttaacatccctctgataaactaaag 313  
|||||  
Db 168213 CTCTTACACGCTCTCCGATTAATAAGG 168242  
|||||

Search completed: September 25, 2001, 20:35:12  
Job time: 22075 sec



QY	104	taaaagatgctgtcaaaaactttatgtgatgtgaacccaagactctgaaagaagaaagagc	163
Db	16393	TATAGATGCTGTGTGAAAACTTTTATGATATAGACACGAGACTCTGAAAGAAAGAGAGC	16452
QY	164	ttatttaaacctgactttacatcttgcttcagatattttctcaagcatcctgtgtaaag	223
Db	16453	TTTATTATTAACCTGACCTTTACATCTTGCTGCATTTTCTTCAGATTTTCTCAAGCATTCGTATTAAG	16512
QY	224	atgttcgcttaactgtagtcctgcgcgccttgctgatatatttcaggatttatgcttcgaag	283
Db	16513	ATGTTTCGCTTACCTGGTACCCGTCGCTTGGCTGTGATATTTTCAGGATTTATGCTCTCTTAAG	16572
QY	284	ctcctcacatccctcgtataactaaag	313
Db	16573	CTCCTTACACATCCCTGTATTAACCTAAAGG	16602
RESULT	14		
LOCUS	AL353724	163642 bp	DNA
DEFINITION	Homo sapiens chromosome 13 clone RP11-448113, ***	23-JAN-2001	HTG
ACCESSION	AL353724		PROGRAMS: ***, 9 unordered pieces.
VERSION	AL353724.3	GI:9930876	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 163642)		
AUTHORS	Burton,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk		
	On Aug 27, 2000 this sequence version replaced gi:9926534.		
COMMENT	----- Genome Center		
	Center: Sanger Centre		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: humquerry@sanger.ac.uk		
	----- Project Information		
	Center project name: Dh448113		
	----- Summary Statistics		
	Assembly program: XGAP; version 4.5		
	Sequencing vector: plasmid; L08752; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Consensus quality: 159306 bases at least Q40		
	Consensus quality: 161143 bases at least Q30		
	Consensus quality: 162084 bases at least Q20		
	Insert size: 162842; sum-of-contigs		
	Insert size: 163676; 2.5% error; agarose-fp		
	Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality		
	coverage: 4.42x in Q20 bases; agarose-fp		
	-----		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 9 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	* 1 20261: contig of 20261 bp in length		
	* 20262 20361: gap of 100 bp		
	* 20362 38140: contig of 17779 bp in length		
	* 38141 38240: gap of 100 bp		
	* 38241 85018: contig of 46778 bp in length		
	* 85019 85118: gap of 100 bp		
	* 85119 93385: contig of 8267 bp in length		
	* 93386 93485: gap of 100 bp		
	* 93486 96470: contig of 2985 bp in length		

FEATURES	SOURCE	Location/Qualifiers
*	96471	96570: gap of 100 bp
*	96571	137301: contig of 40731 bp in length
*	137302	137401: gap of 100 bp
*	137402	147035: contig of 9634 bp in length
*	147036	147135: gap of 100 bp
*	147136	149710: contig of 2575 bp in length
*	149711	149810: gap of 100 bp
*	149811	163642: contig of 13832 bp in length.
FEATURES		
SOURCE		
misc_feature	1. 163642	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="13"
		/clone="RP11-44813"
		/clone_lib="RPC1-11.2"
misc_feature	1. 20261	/note="assembly_fragment:00330"
		clone_end:77
		vector_side:left
misc_feature	20362. 38140	/note="assembly_fragment:00335"
		fragment_chain:1
misc_feature	38241. 85018	/note="assembly_fragment:01397"
		fragment_chain:1
misc_feature	85119. 93385	/note="assembly_fragment:01083"
		fragment_chain:1
misc_feature	93486. 96470	/note="assembly_fragment:00197"
		fragment_chain:1
misc_feature	96571. 137301	/note="assembly_fragment:00284"
		fragment_chain:2
misc_feature	137402. 147035	/note="assembly_fragment:00874"
		fragment_chain:2
misc_feature	147136. 149710	/note="assembly_fragment:00391"
		fragment_chain:3
misc_feature	149811. 163642	/note="assembly_fragment:00398"
		fragment_chain:3
BASE COUNT	54756 a 31368 c 29374 g 47337 t 807 others	
ORIGIN		
Query Match	5.0%: Score 206.8; DB 79; Length 163642;	
Best Local Similarity	99.0%: Pred No. 2.1e-32;	
Matches 208; Conservative	0; Mismatches 2; Indels 0; Gaps	
QY	104	taaaagatggtgtgtaaaactttatagatagatagaccagagactctgaagaagaagaagagc 163
Db	32941	TATAGATGGTGTGAAAACTTTATAGGATATGAGCAGACGACTGTGAAGAAAGAGC 32882
QY	164	ttatttaacctagcttaacctctgtctcagattttttctcaagcctctggaag 223
Db	32881	TTTATTTAAACCTACCTTTATCAATCTTGCTTCAGATTTTTCACAGCATCTGATTAAG 32822
QY	224	atgttcgctactgtgtaagcctctcgtcctctgcgataatlttcaggaattatgtcctgaag 283
Db	32821	ATGTTGCGTACTATGGAGCGCTGCTGCTTCGATATTTTCAGAGATTTATGTCCTCGAAG 32766
QY	284	ctccttacacatccctctgataactaaag 313
Db	32761	CTCCTTACACATCCCTGATTAACATAAGG 32732
RESULT 15		
LOCUS	AL358892	195380 bp DNA HTG 06-MAR-2001
DEFINITION	Mus musculus chromosome 5 clone RP21-41766, *** SEQUENCING IN	
	PROGRESS ***, in unordered pieces.	



```
repeat_region 9724. .9876
/partial
/note="Alu repeat: matches 158. .1 of consensus"
repeat_region 10567. .10594
/note="14 copies of 2 mer 93 & conserved"
repeat_region 10616. .10900
/partial
/note="Alu repeat: matches 300. .1 of consensus"
repeat_region 10960. .10985
/note="13 copies of 2 mer 92 & conserved"
repeat_region 11912. .12100
/partial
/note="Alu repeat: matches 268. .65 of consensus"
repeat_region 17050. .17359
/note="Alu repeat: matches 1. .306 of consensus"
repeat_region 17639. .17888
/partial
/note="Alu repeat: matches 34. .306 of consensus"
repeat_region 19151. .19980
/note="11 element fragment"
repeat_region 19981. .20264
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 20238. .21190
/note="11 element fragment"
repeat_region 21199. .21488
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 21538. .22226
/note="11 element fragment"
repeat_region 25696. .25989
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 26902. .27167
/note="2 copies of 133 mer 93 & conserved"
repeat_region 27850. .27930
/note="11 element fragment"
repeat_region 28002. .28043
/note="11 element fragment"
repeat_region 28333. .28780
/note="11 element fragment"
repeat_region 29176. .29449
/note="Alu repeat: matches 1. .306 of consensus"
repeat_region 29904. .30192
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 30354. .30663
/note="11 element fragment"
repeat_region 30683. .30972
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 30977. .31204
/note="11 element fragment"
repeat_region 34819. .35121
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 37059. .37282
/note="MIR2B2 element fragment"
repeat_region 37453. .37506
/note="MIR2B2 element fragment"
repeat_region 38414. .38706
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 39516. .39563
/note="24 copies of 2 mer 81 & conserved"
repeat_region 41244. .41535
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 43994. .44242
/partial
/note="Alu repeat: matches 44. .308 of consensus"
complement(44369. .44677)
/note="match: 217151 DNA segment containing (CA) repeat"
complement(44479. .44510)
/note="16 copies of CA 100% conserved; Polymorphic by
comparison with 217151"

repeat_region 44702. .44889
/note="2 copies of 94 mer 85 & conserved"
repeat_region 45935. .46229
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 46910. .46988
/note="11 element fragment"
repeat_region 47081. .47119
/note="11 element fragment"
repeat_region 47168. .47431
/note="11 element fragment"
repeat_region 47615. .47902
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 47922. .48213
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 49426. .49717
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 50248. .50533
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 51097. .51380
/note="Alu repeat: matches 1. .304 of consensus"
repeat_region 62318. .62543
/partial
/note="Alu repeat: matches 302. .66 of consensus"
repeat_region 63358. .63613
/partial
/note="Alu repeat: matches 306. .34 of consensus"
repeat_region 66191. .66236
/note="23 copies of 2 mer 80 & conserved"
repeat_region 68771. .69062
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 70282. .70559
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 70590. .70881
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 75593. .75883
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 75974. .76251
/partial
/note="Alu repeat: matches 294. .1 of consensus"
repeat_region 76531. .76608
/note="MIR element fragment"
repeat_region 78482. .78832
/note="MER11B element fragment"
repeat_region 78574. .78743
/note="MER11A element fragment"
repeat_region 78941. .79021
/note="MER11B element fragment"
repeat_region 79249. .79709
/note="MER11A element fragment"
repeat_region 80233. .80302
/note="SVA element fragment"
repeat_region 80355. .80535
/partial
/note="Alu repeat: matches 213. .20 of consensus"
repeat_region 80563. .80619
/partial
/note="Alu repeat: matches 277. .221 of consensus"
repeat_region 80651. .82294
/note="SVA element fragment"
repeat_region 82753. .83043
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 83149. .83296

Query Match 5.0%; Score 206.8; DB 92; Length 113704;
Best Local Similarity 99.0%; Pred. No. 2.1e-32;
Matches 208; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```



29585	30328:	contly of 744 bp	in length
30329	30428:	gap of 100 bp	
30429	31111:	contly of 683 bp	in length
31112	31211:	gap of 100 bp	
31212	31931:	contly of 720 bp	in length
31932	32031:	gap of 100 bp	
32032	32710:	contly of 679 bp	in length
32711	32810:	gap of 100 bp	
32811	33502:	contly of 692 bp	in length
33503	33602:	gap of 100 bp	
33603	34301:	contly of 699 bp	in length
34302	34401:	gap of 100 bp	
34402	35100:	contly of 699 bp	in length
35101	35200:	gap of 100 bp	
35201	35914:	contly of 714 bp	in length
35915	36014:	gap of 100 bp	
36015	36709:	contly of 695 bp	in length
36710	36809:	gap of 100 bp	
36810	37518:	contly of 709 bp	in length
37519	37618:	gap of 100 bp	
37619	38330:	contly of 712 bp	in length
38331	38430:	gap of 100 bp	
38431	39109:	contly of 679 bp	in length
39110	39209:	gap of 100 bp	
39210	39912:	contly of 703 bp	in length
39913	40012:	gap of 100 bp	
40013	40712:	contly of 700 bp	in length
40713	40812:	gap of 100 bp	
40813	41514:	contly of 702 bp	in length
41515	41614:	gap of 100 bp	
41615	42294:	contly of 680 bp	in length
42295	42394:	gap of 100 bp	
42395	43094:	contly of 700 bp	in length
43095	43194:	gap of 100 bp	
43195	43915:	contly of 721 bp	in length
43916	44015:	gap of 100 bp	
44016	44740:	contly of 725 bp	in length
44741	44840:	gap of 100 bp	
44841	45533:	contly of 693 bp	in length
45534	45633:	gap of 100 bp	
45634	46330:	contly of 697 bp	in length
46331	46430:	gap of 100 bp	
46431	47137:	contly of 707 bp	in length
47138	47237:	gap of 100 bp	
47238	47927:	contly of 690 bp	in length
47928	48027:	gap of 100 bp	
48028	48709:	contly of 682 bp	in length
48710	48809:	gap of 100 bp	
48810	49509:	contly of 700 bp	in length
49510	49609:	gap of 100 bp	
49610	50331:	contly of 722 bp	in length
50332	50431:	gap of 100 bp	
50432	51135:	contly of 704 bp	in length
51136	51235:	gap of 100 bp	
51236	51981:	contly of 746 bp	in length
51982	52081:	gap of 100 bp	
52082	52783:	contly of 702 bp	in length
52784	52883:	gap of 100 bp	
52884	53580:	contly of 697 bp	in length
53581	53680:	gap of 100 bp	
53681	54398:	contly of 718 bp	in length.

Oy	2789	tttcgcgttaagagcttccactgagtataigcaatcgtgaccttggcgaaaagtc	2848
Db	4472	ttttccctttttagcgccttccactgttaataatggaacattctgaccttggcgaaaagtc	4531
Oy	2849	ctctgaagagagagaagagcctatgcctctagggcaatgcttggcgaaaataataatgtaagc	2908
Db	4532	ctgtgaagagagagaagagcctatgcctctagggcaatgcttggcgaaaataataatgtaagc	4591
Oy	2909	ggagagtactgaagcaagcatgcagctgctttagtgaana	2945
Db	4592	ggagagtactgaagcaagcatgcagctgctttagtgaana	4628

RESULT	13
LOCUS	HS267P19
DEFINITION	Human DNA sequence from cosmid 267P19, BRCA2 gene region chromosome
ACCESSION	U08001
VERSION	275889
KEYWORDS	13q12-13; repeat polymorphism.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	Odell.C.
JOURNAL	Direct Submission Submitted (04-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,

COMMENT

This sequence is not the entire insert of clone 267P19. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 267P19 is at 1 in this sequence. The true right end of clone 26H23 is at 24539.

The true left end of clone 49J10 is at 113606.

267P19 is from the human PAC library.

FEATURES	Location/Qualifiers
source	1. .113704

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-13"
/clone="XX-267p19"
1..235
/partial
/note="Alu repeat: matches 252. .1 of consensus"
370..662
/partial
/note="Alu repeat: matches 308. .1 of consensus"
4849..5140
/partial
/note="Alu repeat: matches 308. .1 of consensus"
5367..5660
/note="Alu repeat: matches 1. .308 of consensus"
6080..6187
/note="2 copies of 54 mer 87 % conserved"
8133..8401
/note="MER7A element fragment"
8623..8748
/note="L1 element fragment"
9617..9711
/partial
/note="Alu repeat: matches 305. .210 of consensus"

```



Db	481	GAAGGAACACCTAGCTTAGAAGACACTGGCACTGAGACTGGAGCAATATTATGTAATTT	540
Oy	3485	caaatcaagctctcctggaagataaagggaagcttgaatgtctga	3533
Db	541	CAGAGCTGAACCCCTTCAACCGAATTCGATCAACGGACAGATTTCAGA	589
RESULT	12		
LOCUS	AC068224		
DEFINITION	AC068224	54398 bp DNA HTG	30-APR-2000
ACCESSION	AC068224	Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS	
VERSION	AC068224.1	GI:7671284	
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 54398)		
JOURNAL	Bioren, B., Linton, L., Nusbaum, C. and Lander, E.		
AUTHORS	Homo sapiens chromosome 3, clone RP11-660H19		
COMMENT	Unpublished		
	2 (bases 1 to 54398)		
	Bioren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,		
	Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,		
	Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,		
	Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,		
	Collamore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S.,		
	Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,		
	Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,		
	Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,		
	Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,		
	Klein, J., Larocque, K., Lamazares, R., Landers, T., Lechoczky, J.,		
	Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Marguis, N.,		
	McCarthy, M., McKean, P., McGuirk, A., McKernan, K., McPheters, R.,		
	Meldrum, J., Menues, L., Mihova, T., Miranda, C., Mleaga, V., Morrow, J.		
	Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,		
	O'Neill, D., Olyvar, T.M., Oliver, J., Peterson, K., Pierre, N.,		
	Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,		
	Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,		
	Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,		
	Tesfaye, S., Theodore, J., Tirelli, A., Travers, M., Triggillo, U.,		
	Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,		
	Young, G., Zainoun, J., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome		
COMMENT	Research, 330 Charles Street, Cambridge, MA 02141, USA		
	All repeats were identified using RepeatMasker:		
	Smith, A.F.A. & Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RW/RepeatMasker.html		
	Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence_submissions@genome.wi.mit.edu		
	Project Information		
	Center project name: L10161		
	Center clone name: 660_H_19		
	NOTE: This record contains 68 individual		
	* sequencing reads that have not been assembled into		
	* contigs. Runs of N are used to separate the reads		
	* and the order in which they appear is completely		
	* arbitrary. Low-pass sequence sampling is useful for		
	* identifying clones that may be gene-rich and allows		
	* overlap relationships among clones to be deduced.		
	* However, it should not be assumed that this clone		
	* will be sequenced to completion. In the event that		
	* the record is updated, the accession number will		
	* be preserved.		
	1 685: contig of 685 bp in length		

[illegible]







## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence-submissions@genome.wi.mit.edu](mailto:sequence-submissions@genome.wi.mit.edu)

## ----- Project Information

Center project name: L10161

Center clone name: 660\_H\_19

\* NOTE: This record contains 68 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 685: contig of 685 bp in length  
\* 686 785: gap of 100 bp  
\* 786 1513: contig of 728 bp in length  
\* 1514 1613: gap of 100 bp  
\* 1614 2354: contig of 741 bp in length  
\* 2355 2454: gap of 100 bp  
\* 2455 3184: contig of 730 bp in length  
\* 3185 3284: gap of 100 bp  
\* 3285 3959: contig of 675 bp in length  
\* 3960 4059: gap of 100 bp  
\* 4060 4746: contig of 687 bp in length  
\* 4747 4846: gap of 100 bp  
\* 4847 5564: contig of 718 bp in length  
\* 5565 5664: gap of 100 bp  
\* 5665 6360: contig of 696 bp in length  
\* 6361 6460: gap of 100 bp  
\* 6461 7132: contig of 672 bp in length  
\* 7133 7232: gap of 100 bp  
\* 7233 7937: contig of 705 bp in length  
\* 7938 8037: gap of 100 bp  
\* 8038 8749: contig of 712 bp in length  
\* 8750 8849: gap of 100 bp  
\* 8850 9522: contig of 673 bp in length  
\* 9523 9622: gap of 100 bp  
\* 9623 10356: contig of 734 bp in length  
\* 10357 10456: gap of 100 bp  
\* 10457 11181: contig of 725 bp in length  
\* 11182 11281: gap of 100 bp  
\* 11282 11975: contig of 694 bp in length  
\* 11976 12075: gap of 100 bp  
\* 12076 12785: contig of 710 bp in length  
\* 12786 12885: gap of 100 bp  
\* 12886 13557: contig of 672 bp in length  
\* 13558 13657: gap of 100 bp  
\* 13658 14347: contig of 690 bp in length  
\* 14348 14447: gap of 100 bp  
\* 14448 15132: contig of 685 bp in length  
\* 15133 15232: gap of 100 bp  
\* 15233 15912: contig of 680 bp in length  
\* 15913 16012: gap of 100 bp  
\* 16013 16731: contig of 719 bp in length  
\* 16732 16831: gap of 100 bp  
\* 16832 17506: contig of 675 bp in length  
\* 17507 17606: gap of 100 bp  
\* 17607 18339: contig of 733 bp in length  
\* 18340 18439: gap of 100 bp  
\* 18440 19109: contig of 670 bp in length  
\* 19110 19209: gap of 100 bp  
\* 19210 19900: contig of 691 bp in length  
\* 19901 20000: gap of 100 bp

\* 20001 20693: contig of 693 bp in length  
\* 20694 20793: gap of 100 bp  
\* 20794 21464: contig of 671 bp in length  
\* 21465 21564: gap of 100 bp  
\* 21565 22245: contig of 681 bp in length  
\* 22246 22345: gap of 100 bp  
\* 22346 23049: contig of 704 bp in length  
\* 23050 23149: gap of 100 bp  
\* 23150 23857: contig of 708 bp in length  
\* 23858 23957: gap of 100 bp  
\* 23958 24685: contig of 728 bp in length  
\* 24686 24785: gap of 100 bp  
\* 24786 25499: contig of 714 bp in length  
\* 25500 25599: gap of 100 bp  
\* 25600 26302: contig of 703 bp in length  
\* 26303 26402: gap of 100 bp  
\* 26403 27108: contig of 706 bp in length  
\* 27109 27208: gap of 100 bp  
\* 27209 27910: contig of 702 bp in length  
\* 27911 28010: gap of 100 bp  
\* 28011 28720: contig of 710 bp in length  
\* 28721 28820: gap of 100 bp  
\* 28821 29484: contig of 664 bp in length  
\* 29485 29584: gap of 100 bp  
\* 29585 30328: contig of 744 bp in length  
\* 30329 30428: gap of 100 bp  
\* 30429 31111: contig of 683 bp in length  
\* 31112 31211: gap of 100 bp  
\* 31212 31931: contig of 720 bp in length  
\* 31932 32031: gap of 100 bp  
\* 32032 32710: contig of 679 bp in length  
\* 32711 32810: gap of 100 bp  
\* 32811 33502: contig of 692 bp in length  
\* 33503 33602: gap of 100 bp  
\* 33603 34301: contig of 699 bp in length  
\* 34302 34401: gap of 100 bp  
\* 34402 35100: contig of 699 bp in length  
\* 35101 35200: gap of 100 bp  
\* 35201 35914: contig of 714 bp in length  
\* 35915 36014: gap of 100 bp  
\* 36015 36709: contig of 695 bp in length  
\* 36710 36809: gap of 100 bp  
\* 36810 37518: contig of 709 bp in length  
\* 37519 37618: gap of 100 bp  
\* 37619 38330: contig of 712 bp in length  
\* 38331 38430: gap of 100 bp  
\* 38431 39109: contig of 679 bp in length  
\* 39110 39209: gap of 100 bp  
\* 39210 39912: contig of 703 bp in length  
\* 39913 40012: gap of 100 bp  
\* 40013 40712: contig of 700 bp in length  
\* 40713 40812: gap of 100 bp  
\* 40813 41514: contig of 702 bp in length  
\* 41515 41614: gap of 100 bp  
\* 41615 42294: contig of 680 bp in length  
\* 42295 42394: gap of 100 bp  
\* 42395 43094: contig of 700 bp in length  
\* 43095 43194: gap of 100 bp  
\* 43195 43915: contig of 721 bp in length  
\* 43916 44015: gap of 100 bp  
\* 44016 44740: contig of 725 bp in length  
\* 44741 44840: gap of 100 bp  
\* 44841 45533: contig of 693 bp in length  
\* 45534 45633: gap of 100 bp  
\* 45634 46330: contig of 697 bp in length  
\* 46331 46430: gap of 100 bp  
\* 46431 47137: contig of 707 bp in length  
\* 47138 47237: gap of 100 bp  
\* 47238 47927: contig of 690 bp in length  
\* 47928 48027: gap of 100 bp  
\* 48028 48709: contig of 682 bp in length  
\* 48710 48809: gap of 100 bp  
\* 48810 49509: contig of 700 bp in length







OY	4071	atcttcgtgatctagtgcgaattgaaatccacacagtcaccacacagaagaagcgaggag	4130
Db	21133	ATCTTCGTGAATCTAGTGCGAATTGAAATCCACACAGTCCACACCACAGAAGAAGCGAGGAG	21192
OY	4131	acctcaaaaagccatcacccatcatcacaacaaaaaatgfcg	4173
Db	21193	ACCAATCAAAAACGCCATCATCACATCAACAACAAAAAATAATGTC	21235
RESULT	9		
LOCUS	AC016449/c		
DEFINITION	Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.		
ACCESSION	AC016449		
VERSION	AC016449.2 GI:9104517		
KEYWORDS	HTGS_PHAMSE0.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 72157)		
JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
AUTHORS	Homo sapiens, clone RP11-45L14		
REFERENCE	Unpublished		
TITLE	2 (bases 1 to 72157)		
JOURNAL	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,		
AUTHORS	Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,		
REFERENCE	Cooke,P., Castle,A., Colangelo,M., Collins,S., Collamore,A.,		
AUTHORS	Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,		
REFERENCE	Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,		
AUTHORS	Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,		
REFERENCE	Leloczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,		
AUTHORS	McKean,P., McGurt,A., McKernan,K., McLaughlin,J., Meldrim,J.,		
REFERENCE	Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,		
AUTHORS	Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,		
REFERENCE	Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,		
AUTHORS	Testafay,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,		
REFERENCE	Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome		
AUTHORS	Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	On Jul 13, 2000 this sequence version replaced gi:6479175.		
AUTHORS	All repeats were identified using RepeatMasker:		
COMMENT	Smit, A.F.A. & Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RM/repeatmasker.html		
	----- Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIBR		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence.submissions@genome.wi.mit.edu		
	----- Project Information		
	Center project name: L1189		
	Center clone name: 45_L_14		
	-----		
*	* NOTE: This record contains 83 individual		
*	* sequencing reads that have not been assembled into		
*	* contigs. Runs of N are used to separate the reads		
*	* and the order in which they appear is completely		
*	* arbitrary. Low-pass sequence sampling is useful for		
*	* identifying clones that may be gene-rich and allows		
*	* overlap relationships among clones to be deduced.		
*	* However, it should not be assumed that this clone		
*	* will be sequenced to completion. In the event that		
*	* the record is updated, the accession number will		
*	* be preserved.		
*	1 757: contig of 757 bp in length		
*	758 857: gap of 100 bp		
*	858 1635: contig of 778 bp in length		
*	1636 1735: gap of 100 bp		
*	1736 2490: contig of 755 bp in length		

[illegible]







repeat\_region  
5620..5810  
/note="MIR repeat: matches 58..262 of consensus"  
7829..8133  
/note="AluX repeat: matches 1..301 of consensus"  
repeat\_region  
10957..11242  
/note="AluY repeat: matches 291..1 of consensus"  
repeat\_region  
12344..12405  
/note="L1PA16 repeat: matches 903..842 of consensus"  
repeat\_region  
12381..13119  
/note="L1PA16 repeat: matches 833..68 of consensus"  
repeat\_region  
14428..14540  
/note="HY1 repeat: matches 110..1 of consensus"  
repeat\_region  
14951..15071  
/note="MIR2 repeat: matches 126..1 of consensus"  
repeat\_region  
16589..16885  
/note="AluX repeat: matches 302..4 of consensus"  
repeat\_region  
18067..18397  
/note="L1 repeat: matches 3807..4127 of consensus"  
repeat\_region  
19076..19481  
/note="MLT1A2 repeat: matches 374..1 of consensus"  
repeat\_region  
19516..19799  
/note="AluX repeat: matches 19..302 of consensus;  
incomplete repeat"  
repeat\_region  
19808..20127  
/note="AluY repeat: matches 2..299 of consensus"  
prim\_transcript  
21359..26899  
/note="match: multiple ESTs; match: H44781 D20082 W20070  
T64867 H45693; match: F12354 N91467 N83765 H45741 H93424"  
repeat\_region  
24795..25089  
/note="AluX repeat: matches 294..1 of consensus"  
repeat\_region  
26083..26106  
/note="L12 copies of 2 mer 96 & conserved"  
prim\_transcript  
complement(27120..27476)  
/note="match: multiple ESTs; match F09972 T72607"  
prim\_transcript  
227633..28101  
/note="match: multiple ESTs; match: 5' EST T51695 clone  
72405; Paired with EST T51543 matching this clone; match  
H91918"  
repeat\_region  
28327..28360  
/note="17 copies of 2 mer 88 & conserved"  
prim\_transcript  
228864..29013  
/note="match: 3' EST T51543 clone 72405; Paired with EST  
T51695 matching this clone"  
repeat\_region  
29300..29456  
/note="MIR2 repeat: matches 146..1 of consensus"  
repeat\_region  
30648..30934  
/note="AluX repeat: matches 7..302 of consensus"  
repeat\_region  
33156..33439  
/note="AluX repeat: matches 301..2 of consensus"  
repeat\_region  
34900..35062  
/note="L1MD2 repeat: matches 425..263 of consensus"  
repeat\_region  
34982..35061  
/note="L1ME3 repeat: matches 343..264 of consensus"  
repeat\_region  
37923..38245  
/note="AluX repeat: matches 294..1 of consensus"  
repeat\_region  
38457..38556  
/note="MIR repeat: matches 61..161 of consensus"  
repeat\_region  
39265..39568  
/note="AluO repeat: matches 302..3 of consensus"  
repeat\_region  
40952..41046  
/note="MIR repeat: matches 29..125 of consensus"  
repeat\_region  
43042..43331  
/note="AluO repeat: matches 301..11 of consensus"  
repeat\_region  
43583..43669  
/note="MERSA repeat: matches 10..99 of consensus"  
repeat\_region  
43881..44083  
/note="MIR repeat: matches 248..28 of consensus"  
repeat\_region  
44772..45170  
/note="MIR repeat: matches 388..1 of consensus"  
repeat\_region  
45848..45949  
/note="MIR repeat: matches 144..41 of consensus"  
repeat\_region  
46073..46373

repeat\_region  
49357..49436  
/note="AluSp repeat: matches 298..1 of consensus"  
repeat\_region  
50023..50324  
/note="MIR repeat: matches 198..114 of consensus"  
repeat\_region  
50826..50892  
/note="AluX repeat: matches 1..302 of consensus"  
repeat\_region  
51407..51534  
/note="L1MB8 repeat: matches 853..921 of consensus"  
repeat\_region  
51563..51863  
/note="L1 repeat: matches 4869..4999 of consensus"  
repeat\_region  
51869..52077  
/note="AluY repeat: matches 1..301 of consensus"  
repeat\_region  
52311..52712  
/note="AluO repeat: matches 86..298 of consensus;  
incomplete repeat"  
repeat\_region  
52698..52781  
/note="L1ME3A repeat: matches 44..475 of consensus"  
repeat\_region  
52719..52833  
/note="L1MD2 repeat: matches 569..653 of consensus"  
repeat\_region  
53752..54490  
/note="L1MB3 repeat: matches 591..707 of consensus"  
repeat\_region  
54313..54870  
/note="L1 repeat: matches 4668..5390 of consensus"  
repeat\_region  
55008..55520  
/note="L1MB4 repeat: matches 1..540 of consensus"  
repeat\_region  
57267..57487  
/note="L1MA9 repeat: matches 530..1055 of consensus"  
repeat\_region  
57510..57576  
/note="AluY repeat: matches 79..298 of consensus;  
incomplete repeat"  
repeat\_region  
58670..58699  
/note="MIR repeat: matches 76..143 of consensus"  
repeat\_region  
58676..58699  
/note="L15 copies of 2 mer 87 & conserved"  
repeat\_region  
59808..60115  
/note="6 copies of 4 mer 100 & conserved"  
repeat\_region  
60135..60526  
/note="L1MB7 repeat: matches 895..571 of consensus"  
repeat\_region  
60531..60815  
/note="MIR1D repeat: matches 505..108 of consensus"  
repeat\_region  
60816..60915  
/note="AluO repeat: matches 2..290 of consensus"  
repeat\_region  
60918..61113  
/note="MIR1D repeat: matches 114..3 of consensus"  
repeat\_region  
61115..61379  
/note="L1MB6 repeat: matches 529..320 of consensus"  
repeat\_region  
61380..61675  
/note="AluY repeat: matches 299..31 of consensus;  
incomplete repeat"  
repeat\_region  
61540..61805  
/note="L1MB5 repeat: matches 309..1 of consensus"  
repeat\_region  
61916..62212  
/note="L1 repeat: matches 5389..5107 of consensus"  
repeat\_region  
64361..64420  
/note="AluSc repeat: matches 1..299 of consensus"  
repeat\_region  
64816..65101  
/note="U6 repeat: matches 63..1 of consensus"  
repeat\_region  
65202..65856  
/note="AluO repeat: matches 2..302 of consensus"  
repeat\_region  
66393..66470  
/note="L1PA16 repeat: matches 902..217 of consensus"  
repeat\_region  
66530..66868  
/note="L1 repeat: matches 2101..2178 of consensus"  
repeat\_region  
66881..67104  
/note="L1 repeat: matches 2265..2610 of consensus"  
repeat\_region  
67104..67855  
/note="MIR30 repeat: matches 2..230 of consensus"  
repeat\_region  
67999..68677  
/note="L1 repeat: matches 2613..3397 of consensus"  
repeat\_region  
68693..68972  
/note="L1 repeat: matches 4255..4942 of consensus"  
repeat\_region  
68983..69436  
/note="AluSc repeat: matches 299..1 of consensus"







Db 121 CTACATCAGAGGCTACTGTTCTGCCATGTTTGGAAAACTGATGACCATACCAAGAAATT 180  
 QY 1607 taccgtacccgtgtaaggtcgaagatttcagtaagaatcacacaggtgttaagaatg 1666  
 Db 181 TGCCGTGACCCCGGAAAGCAGATTTTGTGAAGAAATTTAAACAGGTTCTCCGCGATG 240  
 QY 1667 atgagaataaagaacagtlagaagtlactgttlaagtcacacatgctcctcgtcagag 1726  
 Db 241 ATGAGAAATCTGGCTCTCAGTTGAGTTATTAATTAGCCCAACCTGTTCTTGAACAAAG 300  
 QY 1727 ctgaaggtgtgtcgtggaataactaagaagttggcgaaccccaacagctcaaatc 1786  
 Db 301 CAGTAATTTGTGTAGAGAAATTAAGCCGAAACTTGGCAAAATCTTAAGCAACCAATTC 360  
 QY 1787 ctcccggaatgatcaagttctctctggaagatagacacctgtgacatagatccg 1846  
 Db 361 CTTTTCAGAGATGCTCAAAATTTCTGTGGAAATGCGACCTGTGACACTTGTATTCAG 420  
 QY 1847 aatcatalcagtlccttalttaacaagtgaaacaatatagatgaaacagcagatgag 1906  
 Db 421 AAGCATTAAGTGCACAGTGAATTTGATGAATTAAGTCAATAGAGGGGACAGCATGTATG 480  
 QY 1907 aagatgggtgtgtcccaatgataagcaccagagctcctgaactgttaagtac 1966  
 Db 481 AAGAGGAGGAGTGTAGTCCAGATACAGCTATCCGTTCAAGACTTGAACCTTCTTAAGTTC 540  
 QY 1967 tctattacacatcccatccatcattcattcgtctgtaaacattgaaatcattactgctt 2026  
 Db 541 TGCTTTTACACATCTTACCTTCGTTCCACTGTGCAAGACATATGAGTCTTGTATTCAGT 600  
 QY 2027 gctcgaataatgaaagaaagtagaagaagctgacatacaatttcaacaacagac 2086  
 Db 601 GCCTAAATAATGAGATGACAGAGTAGAGAGAGCTGATTAACAATTTTGAATAATACAG 660  
 QY 2087 gaagcaaatatgaagaagatttccacacatcagatcagacctgtcttccgttttacc 2146  
 Db 661 GTCAAAATAATGAAACAGACCTCCCGAGATGATGATGATGATGATGATGATGATGATG 720  
 QY 2147 acaaatcaaaaagaagcccccgcaagcaaatatgcatcattcattcattcattcattc 2206  
 Db 721 AAAAAGCAAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
 QY 2207 taattctagtaagaagaccagttgcacagataattgagcctctcgtatagaagcctag 2266  
 Db 781 TATTCACAAATAATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
 QY 2267 atccaaagaacctggaacatctcaacaacatggttactatggtatcattgtctcc 2326  
 Db 841 ATGCTGATGTCAGAAACAACTTAACTCCATTAGTTTCAATGGGCCCAATTTCTATGT 900  
 QY 2327 ttgcaactgatacaattgctgtccttggaaatcttggtagtacttacttacttacttact 2386  
 Db 901 TAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
 QY 2387 atcttcatgaatgatcgtgcttccagggaagaagaacaactaaacttgggttccagag 2446  
 Db 961 ATCTGTAATGAATGACAGTCAACAGGTGAAGAAATGAAAGAACTGCTCTCCAGATG 1020  
 QY 2447 aagaagatctccctggaagaacatgltcaaaatcagcgtcattaaatgagtgatgag 2506  
 Db 1021 AAGAGGTTTCCCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
 QY 2507 tacttgaatgaaataatcaatcacagtaaatcaggaacttccacttaagatgcttaacaa 2566  
 Db 1081 TGTTGGGTATGAAAACACAGCTTAATCTGCCAATTTCAACCTTGGTTATTTATAG 1140  
 QY 2567 caatattgatgtatgagagacttgaacagacaggggaaataatgataaacagatag 2626  
 Db 1141 CGAGTGTGTTAGTGAAGGTGACCTGACAGAGCAAAAGAGATGATGATGATGATGATGATG 1200  
 QY 2627 caagtcgagagactgctgctggaggtgtctatgtgaaagctgacaaagaccgttacc 2686  
 Db 1201 CTCGCTTGGCATTTAGTGTGCTGAGTGCATTAATGAAGCTTGTCAAGAACCTTGTATAC 1260

QY 2687 atgaatcatcacatitgaacaatalcagctatgtcattagatcacaagatgaatgct 2746  
 Db 1261 ATGAAATATTATACCCAGCAACAGTTTACGCTCTGTGCACTGTATTAATGATGATGATG 1320  
 QY 2747 atcaagtaagaagtglttgcgaagaacttcaacaagaccttccgtttagcgttc 2806  
 Db 1321 ACCAAGTAAAGCAGATATTTCTCAGAGCTGCATTAAGGACACTTGTGAAGTTACTGCTCC 1380  
 QY 2807 caacttgaatagtgcaatctgtgcttctgtgcaaaaagatcccgtaagaagaagag 2866  
 Db 1381 CATTGAGATATATGCGATCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 1440  
 QY 2867 ctcatctgaagcaatgttgggtgaanaataatgtaagcggagatcgtgaagcagc 2926  
 Db 1441 CACACGACGACATGTTTACTGAAATATACAGATATGCAAGGAAATACATTAAGCAGA 1500  
 QY 2927 atgaagctgttgaagaataatctgtccttctcaccagatgltgttccataca 2986  
 Db 1501 ATCCATATGCTTACAGAAATTAATATATCATGTTGCTGATATGATGATGATGATGATG 1560  
 QY 2987 ttcaccttggcacaatgaccagatattgtcaagaatcagatattgaacaacttaag 3046  
 Db 1561 TTCACCTGCTAGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
 QY 3047 atgttaagaatgtctgt 3106  
 Db 1621 ATATCAAAAGATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
 QY 3107 gtcacgctttatcagaagaatgltgagaanaalataaacaacaagaatgcccagagac 3166  
 Db 1681 GCCATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740  
 QY 3167 cagatgatacaaaaatgaataaactgtacactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3226  
 Db 1741 CAGATGAATCCAGCAATATTAATAAATGATGATGATGATGATGATGATGATGATGATGATG 1800  
 QY 3227 tcatgcaagaagatcactacatcagtttgaatctcctcaagaccggttactaccagctc 3286  
 Db 1801 TAAATATGTAAGATGCTTGTGCAATGCAAGATGATGATGATGATGATGATGATGATGATG 1860  
 QY 3287 gttcttcaactcaactgcaagaatctcagtaaacacaaataatctcctcctgtaa 3346  
 Db 1861 AATTTTATGACCAACCTGAAAGGACTTCTGTATGATGATGATGATGATGATGATGATGATG 1920  
 QY 3347 tgaatcaatcttcaactcctggaagaacaaacaaacaaacaaacaaacaaacaaacaa 3406  
 Db 1921 CAAAGTACTTCTGTTAAACAGGAAGCCAAAGCCGTGAGTACTAGTCCAGTAATTA 1980  
 QY 3407 agccacttctacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3466  
 Db 1981 AGCCTTATGACGAAGGGAAGGAAGAACCTTATGTAAGAGCACTGCACTGCACTGCA 2040  
 QY 3467 gaaatgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3526  
 Db 2041 GCAATATTAATGATTAATTCAGAGCTGAACCTTCAACCGAATATGATGATGATGATGATG 2100  
 QY 3527 gttctgaatgatcacagtgagaatgaga 3557  
 Db 2101 GTTCAGAGGACAGCAAAACTGAGTTAGTGA 2131

RESULT 6  
 LOCUS AK026889 2079 bp mRNA PRI 29-SEP-2000  
 DEFINITION Homo sapiens cDNA: FLJ23236 fls, clone COL00725.  
 ACCESSION AK026889  
 VERSION AK026889.1 GI:10439854  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL00725.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.



```

|||||
Db 841 ATGAAGATGAGGGTGTTCACATGATCAAGCCATCAGACGAGTCTTGAAAGCTTAAAG 900
Oy 1964 taccctctttacacatccatctcattctcctggaacatttgatattacacg 2023
Db 901 TACTCTCATTTTACATCCATCTCATTTTCATCTGCTGAACATTTGATATCATCTGG 960
Oy 2024 ctgtctctaaatgtgatgtgaaaaagtagcagaagctgcaactacataattccaaaca 2083
Db 961 CTGTGCTCAAAATGATGATGAAAAAGTAGCGAAGCTGCTCATCAAAATTTTCAAAAA 1020
Oy 2084 caggagcaaaaattgaaagagatttccacacatcagatcagccttgcctctgtttac 2143
Db 1021 CAGGAAGCAAAATTTGAAGAGATTTTCCACACATGATGATCAGCTTGTCTCTGTTTAC 1080
Oy 2144 atcaaaaatctaaaaaagaccccccg-ctaaagcacaatatagcattctatctatccat 2202
Db 1081 ATCAACAAATCTMAAAAGAACCCCGCTCAAGCCAAATATGCAATTCATTTGATCCAT 1140
Oy 2203 gcatattctctagtaagaagaccagattgacacagatatattgagcctctgataagagc 2262
Db 1141 GCGATATTTTCTAGTAAGAAGACACAGTTTGACAGATATTTGACCTCTGATTAAGAGC 1200
Oy 2263 ctgagatcaagcaacctggaacatctcacaacacatggttactattgltcatattgct 2322
Db 1201 CTAGATCCAGGACACCTGGAACATCTCATACACCATGTTGTTACTATGTCATATTTGCT 1260
Oy 2323 ctctctgacacctgacatattgctctgctctggaatcttgtagtacttctatctg 2382
Db 1261 CTCTCTGACCTGATCAATTTGCTCTCTCTTGAATCTTTGGTGAATCTTTGATTTG 1320
Oy 2383 aaagatctctcatgaatgatcgcttccagagaaaaagacaactaaacttctggttcca 2442
Db 1321 AAAGATCTTTCATGATGATGATGCTTCCAGGAGAAAAAGACAACTTAATTTGGGTTCCA 1380
Oy 2443 gatgaagaagatctctcctgagaacatggttcaaatcagggctacttaaatgagtgctga 2502
Db 1381 GATGAAGAAGATCTCTGAGACATGCTCAAAATTTCAAGGCTATTAATGATGCTTCCA 1440
Oy 2503 tggctacttggaatgaataaatcaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2561
Db 1441 TGGCTACTTGGATGTATMTTAMTCAAGTATATCAGATCTTCTTCTTAAGSATTTGCT 1500
Oy 2562 aacaacaatctgcatagtagtgaagacttgcacagaacagaggaataatagtaaacaga 2621
Db 1501 AACAAATATTCGATCTGCTGATGAGGGA--TTGMAAGACAGGGGAAA--TTAGTAAACAGA 1557
Oy 2622 tatgtacgctgtagaacttgcctgctgagtgctatctgtgaagctgagcaagaacccctg 2681
Db 1558 TATGTCAAGCTCTGAGACTTGGCTGAGTGTCTATTTRTGAAGCTGGCACAGAACCCCTG 1617
Oy 2682 ttacatataatctacatataatataatataatataatataatataatataatataatataat 2741
Db 1618 TTACATATAATCTCTACATATAATGAACAATATCAGTATGTGCTTATGCTATATCAAGATGA 1677
Oy 2742 atgtcatcaagtaagaacagagtttc-gccacagaacttcacaaagagccttccgcttac 2800
Db 1678 ATGTATCAAGTAAGTAAGAGTGTGCGCCAGAACTTCGCAAGGCTTCCGCTTTCG 1737
Oy 2801 ggtctcaacttgagatataatgcaatctgtgccccttctgtgcaaaagatccctgtaaaagaga 2860
Db 1738 GGTCTTCACTTGAGATATGCAATCTGTGCCCTTTGTGCAAAATATCCTGTMAAGGAGA 1797
Oy 2861 gaagagcctcagtagcagcagctgtgtgtgtaaaataataaagtaagagcagagat 2914
Db 1798 GAAGAGCTCATGCTGAGCAATGTTTGTGTAATAATATTAATTAAGCGGAGCT 1851

```

```

RESULT 5
ABO14548 5177 bp mRNA PRI 06-FEB-1999
LOCUS ABO14548
DEFINITION Homo sapiens mRNA for KIAA0648 protein, partial cds.
ACCESSION ABO14548

```

```

VERSION ABO14548.1 GI:3327109
KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_lib:pbluescriptII
SOURCE SK plus clone:H03994.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 5177)
JOURNAL Ohara, O., Suyama, M., Nagase, T. and Ishikawa, K.
Direct Submission
Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology,
Yara 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
2 (sites)
REFERENCE Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A.,
AUTHORS Kotani, H., Nomura, N. and Ohara, O.
TITLE Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
JOURNAL DNA Res. 5 (3), 169-176 (1998)
FEATURES
source
location/Qualifiers
1..5177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="H03994"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
1..2558
/gene="KIAA0648"
/codon_start=3
/product="KIAA0648 protein"
/protein_id="BAA31623.1"
/db_xref="GI:3327110"
/translation="LYITLASLDNRNAKALNEMKCOMLRSYRELLDLHKOPTSEA
NCSAMFGLMTIANLIDPQKAQDFVKFENVGLDDEKLRQLELLISPTSCROADI
CVREIARKLANPQPTNPFLEMKFLERIPVHIDSEALVALKMKSIETGADDE
EEGVSPDIAISGELKLVLEFTHPTSHSYETSEISLOCRMEDDKVAEAIQFRN
TGHKLENDLPQIRSTLIPILHOKAKRGPHQAKAVHIAHIFNKEVQLQIETPS
RSLNADPQEQIITPLVSLGHSIMLAPDQFAPSPKMSVYANFVYKULMLNDRSTGEKNK
LWSPDEEVSPEVLAKVOAIKILVRLMLGKRNQSKSASVTLRLLSAMLSGDLTEOK
RISKSDMSRLPLAAGSAIMKLAQECVHEIITPQFOLCALVINDECVQVQRIFAOKL
HKAIVKLLPLEYNAIFALCAKDEKERRAHARQCLKNISIRREYIKONPMATEKL
SLPEYVVPYWIHLIADPDPTRSQDVQDLIDRECLFMLEVLMTKNNNSHAFMK
MAENIKILRQNSPDESKTNEKLTVDQVALCVI NSKSAICNADSPKPPVLPKFFPO
PERKQNDKSYISRETRVILLITGKRPACVIGAVNKLPLSAGKRPYRSTGTEGNT
NVNSLENDPSTNRSRBOSSAEAEYGESENEPARIISVTPVKNIIDYKKNKEINSDA
TQGNISSDRGKRTVTYAGAENIQKDEKDESGPPSPSPRRGRPRKPSQGNARK
NDDLNKPLINKRKRRAVGOESPGLGLENAPKRLDIAKKAKAPAEROIQR"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN
Query Match 25.0%; Score 1041.4; DB 85; Length 5177;
Best Local Similarity 68.0%; Pred. No. 4.9e-204;
Matches 1450; Conservative 0; Mismatches 681; Indels 0; Gaps 0;

```



Db 3481 CACCTTTGCCGGGAAAAAGTACAGAGAGACGACTGTGATCTTGTAAAGCTGTAAT 3540  
 QY 3632 tggagaagccctagaagcaaggaataaaacgcccctacagaagcaaggaagaaatlaagta 3691  
 Db 3541 TGGAGAAAGCTTAAGGACGAGAAAAACGCCGTCACAGAACAGAGAGACAAATTAGGTA 3600  
 QY 3692 tggatgacttgaactaagttgltacagaagcaagaaacctaaaggcagtcagcgaagtcgga 3751  
 Db 3601 TGGATGACTTGAAGTATGTTGTCAGAGAACAGAAACTTAAGGAGAGTCAGCAAGTGGGA 3660  
 QY 3752 aaagaagccataagcttcagaatctgataagcaagcagctgctgaggaagaagctca 3811  
 Db 3661 AAAGAGCCCATAGGCTTCAGATCTGATGACAGCATGTCCTGAGGAAAGAGGCTCA 3720  
 QY 3812 aagaagatatattagaataatgaagatgaagacagatagtcgcgaataaagaagtaagaag 3871  
 Db 3721 AAGAAGTATATTAGAAAAAGAGATGACAGATAGTCCGCCAAAAAAGGTTAAAGAG 3780  
 QY 3872 gccgaccacccaacacctctgtgtgaggtacacccaagaagaagccacaatgaacact 3931  
 Db 3781 GCCGACCACCAAAACCTCTGTGTGAGGTACACCAAAAGAGACCMAACATGAAACTT 3840  
 QY 3932 cttaaaaaaggaagcaaaaaaactctggacctccagcagcaaggaaggaagaagaag 3991  
 Db 3841 CTAAAAAAGGACCAAAAAAAATCTGGACCTCAGCAGCAGAGAGAGAGAGAGAGAG 3900  
 QY 3992 aaagaacaagatggaataatagcaagcaagaagtcgaagaagaagcagcagagttcaagga 4051  
 Db 3901 AAGGACAAAGTGAATTCGGAAACAGAGTCCAAAGCAACAGCCAGTGTCAAGGA 3960  
 QY 4052 ggaagcagcagagagagagaatctcctgaatctagtgcaattgaatccacacagtcacac 4111  
 Db 3961 GAGCAGACGAGAGAGAGAGATCTCTGAAATCTAGTGAATTTGAATCCACAGTCCACAC 4020  
 QY 4112 caagaagaaggaaggaaggaagccatcaaaaacgcacatcacatcacacacacacacac 4171  
 Db 4021 CACAGAAAGGACGAGAAAGACCATCAAAAAAGCCATCCATCACCATCAACAAAAAATG 4080  
 QY 4172 t 4172  
 Db 4081 T 4081  
 RESULT 4  
 HSU50533 1852 bp mRNA PRI 27-NOV-1996  
 LOCUS Human BRCA2 region, mRNA sequence CG008.  
 DEFINITION U50533  
 ACCESSION U50533.1 GI:1531605  
 VERSION U50533.1  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 1852)  
 AUTHORS Couch, F.J., Rommens, J.M., Neuhausen, S.L., Belanger, C., Dumont, M., Kenneth, A., Bell, R., Berry, S., Bogen, R., Cannon-Albright, L., Farid, L., Frye, C., Hattler, T., Janacki, T., Jiang, P., Kehrer, R., Leblanc, J., F., McArthur-Morrison, J., McSweney, D., Miki, T., Peng, Y., Samson, C., Schroeder, M., Snyder, S.C., Stringfellow, M., Stroup, C., Swedlund, B., Swensen, J., Teng, D., Thakur, S., Tran, T., Tranchesi, M., Weller-Feldhaus, J., Wong, A.K.C., Shizuya, H., Labrie, F., Skolnick, M.H., Goldgar, D.E., Kamb, A., Weber, B.L., Tavtigian, S.V. and Simard, J.  
 TITLE Generation of an integrated transcription map of the BRCA2 region on chromosome 13q12-q13  
 JOURNAL Genomics 36 (1), 86-99 (1996)  
 MEDLINE 96411650  
 REFERENCE 2 (bases 1 to 1852)  
 AUTHORS Simard, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,

FEATURES Quebec City, Quebec G1V 4G2, Canada  
 source Location/Qualifiers  
 1..1852  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /map="13q12-q13"  
 /note="CG008; DSEG numbers: D13S171 and D13S310"  
 BASE COUNT 609 a 348 c 379 g 505 t 11 others  
 ORIGIN  
 Query Match 42.0%; Score 1754.4; DB 97; Length 1852;  
 Best Local Similarity 98.0%; Pred. No. 0;  
 Matches 1817; Conservative 10; Mismatches 21; Indels 6; Gaps 5;  
 QY 1064 ttaagtgaagtgatcatagaccctgaggaagctattagacatgatatgtgtcaatag 1123  
 Db 1 TTAAGTGAAGTGCACATGACCTGAGGAACTTTACACATGATGTTATTTGTGATAG 60  
 QY 1124 ttacagctgttaaaaggaatattctctgttcaatgatacttaattttgtgagag 1183  
 Db 61 TTACAGCTGTAAAGAGATATTTCTTGTGATATGATCACTTAAATTTGTGAGAG 120  
 QY 1184 agagaacattagacaacagatggaagtaagcaaaagaagcagatggaacttgcccaa 1243  
 Db 121 AGAGAACTTATGACAAACATGATGAGATGAGCAAAAGAGCCATGATGGACTTGCCCAA 180  
 QY 1244 ttataagaatatgctttacagtcagcagctggaagaagatgctgcaaaacagatagat 1303  
 Db 181 TTTATAAGAAATATGCTTTACAGTACACACCTGAAAAAGATGCTGCAAAACATATACAT 240  
 QY 1304 ggaatcaagaacaattgctcatatataatataatcaaaatagattatgatactgactact 1363  
 Db 241 GGATCAAGACAAATTTCTCATATATATATCAAAATAGATTTGATGATCGACTACTTG 300  
 QY 1364 ttgaaggaactctgtcctcaatataatggttcccaaatctagaacatcacagaagatga 1423  
 Db 301 TTGAGCGATCTTGTCTCAATPACATGTTGCTTCACATTTAGAAACATACAGAAAGGATGA 360  
 QY 1424 aatgcttatattactgtatgacacactggaatttaaatgctgtgaaagcaatgaatga 1483  
 Db 361 AATGCTTATATTACTTTATGACCACTGATTTAAATGCTGTGAAAGCATTTGATGAAA 420  
 QY 1484 ttgtgaagatctcaaaactctgctccgaacataaagtaagaagatttgctgactgattaa 1543  
 Db 421 TGTGGAATGTCAAAATCTGCTCGACATCAAGTAAGAAAGATTTGCTTGAATTTAAGC 480  
 QY 1544 aaccacaacagatgcccagtgtaagggcatalttccaagaatgatatattacaagaa 1603  
 Db 481 AACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGATGATGTTATTACAGAA 540  
 QY 1604 attactgatactctgtgaaggtcagaattcatagaagaattcacacagtggttagaag 1663  
 Db 541 ATTACTGTGATCCGCTGTAAGGCTCAGATTTCAAGAAATTCACACAGCTGTTAAAG 600  
 QY 1664 atagatgaagaataagaagaacagttagaagactgtgtatgtaacaaatgctctgcaagc 1723  
 Db 601 ATGTGAGAAATATAGAAAGAGTTAAAGTACTTTGATCCACATGCTCTGCAAGC 660  
 QY 1724 aggtctgaaggtgtgtgctgtaaaactaagaagttggtggcacaaccccaacagcctcaa 1783  
 Db 661 AGGTTGAAGGTGTGTGCTGGAATTAATAAGAAAGTTGGGCAACCCCAAGAGCTTCAA 720  
 QY 1784 atccttctcgtgaagaatgaagttctcttgtagagagataagcactgtgacatagata 1843  
 Db 721 ATCTTCTCTGAATATGATCAAGTTTCTTGGAGAGATAGACCTGTGCACATAGATA 780  
 QY 1844 ccgaatcatcagtgctcttaataaacaagtgaacaacaataatagatggaagcagcaatg 1903  
 Db 781 CCGAATCTATCAGTGTCTTATTAAGAAAGTGAACCAATCAATGATGAGAACAGCAATG 840  
 QY 1904 atgaagatgaggtgtgtccaaactgatcaagccatcagaagcaggtcttgaaactgttaa 1963



1412 cagaacgagatgaatgactatatacttgatgacacacgagatthaagtctgtaag 1471  
1321 CAGAAGGATGAATGCTATATTAATCTGATGCCACATGATTTAAATGCTGTAAG 1380  
1472 ccttgaatgaatgtggaatgttcaaatctgctccgacatacgaagaagattgctg 1531  
1381 CATTGAATGAATGTGGAATGTCAAAATCTCCGACATCAAGTAAGATTTGCTTG 1440  
1532 acttgatgaacacccaacacagatgctcagctcacaatattcaaaagtatg 1591  
1441 ACTTGATTAAGACACCAAAACAGATGCCAGTGTCAAGCCATATTTTCAAAAGTATG 1500  
1592 ttattacaagaattaccgtgacccgtgtaaggtcacaagattcatalgaagaattcacac 1651  
1501 TTATTACAAGAATTATCTGATCTCGTAAGGCTCAGAGATTTCATGAAGAATTACAC 1560  
1652 aggtgttgaagaatgataagaaataagaagaagcagttagaatgactgttgaatccaaat 1711  
1561 AGGTGTGAAGAATGATGGAATAAGAAACAGTAAAGTAAAGTACTTGTAGTCCAAAT 1620  
1712 gctcctgcaagcagctgaaggtgtgtgctggaataaataagaagttggaaccca 1771  
1621 GCTCCTGCAAGCAGGCTGAAGTGTGTGCTCGTAAGTAAGAAAGTTGGGCAACCCCA 1680  
1772 aacagcctacaatcccttccgtgaataatgaatgcttcccttggaagagataagcactg 1831  
1681 AACAGCCTCAAAATCTTCTCGTAAGTATCAAGTTTCTTGGAGAGATAGCACCTG 1740  
1832 tgcacatgatacgaatacactacagtgctcttataaagaatgaaacataatagatg 1891  
1741 TGCAATGATACCCAAATCTATCAGTGCTTTATTAACAAAGTGAACAAATCAATGATG 1800  
1892 gaacagcagatgaatgaatgaatgaggtgtgtccacagatcacaagcacaagcagctgtg 1951  
1801 GAACAGCAGATGATGAAGTGAAGGTTGCCAAGTATCAACCAACACAGCAGGCTTG 1860  
1952 aacgtcctaagtaactcctcattacacatcccatcattcaatcctcgtgaacaattg 2011  
1861 AACTCCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCTGCTGAACATTTG 1920  
2012 aatcattactgctgtcgaataatgagatgataaagaatgagaagtgactacatacaa 2071  
1921 AATCATTAATGCTGTGTGGAATAATGATGAAATAATACAGAAAGTGTGACATCAAAA 1980  
2072 ttctcaaaaacacaggaagcaaatgtgaagagatttccacacatcaatcagccttg 2131  
1981 TTTTCAAAAACACAGGAAGCAAAATTTGAAGAGATTTCACACATCAATCAGCTTGC 2040  
2132 ttcctgtttacatcaacaatctaaanaagaccccccgtaagccaataatgcatc 2191  
2041 TTTCTGTTTACATCACAAAATCTAAAAAAGGACCCCCCGTCAAGCAAAATATGCCATTC 2100  
2192 atgtatcaatgagatattttctagtaagagacccaagtccaagataattgagccctc 2251  
2101 ATTGATCATATGCAATTTCTAGTAAGAGACCCAGTTTCACAGATATTGAGCCCTC 2160  
2252 tgcataaagccttagatccaaagcaacctggaacatctcatacaccaatgttactatg 2311  
2161 TGCATAAGGCTTAGATCAACCAACCTGGAACATCTCTAACAACCATTTGTTACTATTG 2220  
2312 gtcataatgctcctcctgcaactgataatcttctgctgctccttggaatcttggttagcta 2371  
2221 GTCATATTGCTCTCTCTGACCTGATCAATTTGCTGCTCTTTGAAATCTTTGGTAGCTA 2280  
2372 ccttcattgtgaagaatcttctcatgaatgacggtcttcacaggaanaagacaactaac 2431  
2281 CTTTATTGTGAAGAATCTTCTCATGATGATGCTGGCTTCCAGGGAAGAAACAACTAAC 2340  
2432 ttctgggttccagatgaagaagatctcctgtagacaatggttcaaaaattcaggtlatataa 2491  
2341 TTTGGGTTCCAGATGAAGAAGATCTCTCGAGACAAATGTGTAATAATTCAGGCTATTAAAA 2400  
2492 tgatggttcgatgctacttggaatgaaaaataatcaacagtaaatcaaggaacttactc 2551

2401 TCATGTTTCGATGCGCTACTTGAATGAATAAATACACAGTAAATCAGGAACCTTACT 2460  
2552 taagattgtcaacaacaatattgtcagtgatgagagacttgcaaaaagggaataat 2611  
2461 TAAGATTGCTAAACAACAAATTTGATGATGAGACTTGACAAACAGGGGAATAATTA 2520  
2612 gtaaacagataatgcaagctcgaagactgtcgtcggagtgctatctgaaagctgagac 2671  
2521 GTAACCAAGATATGTACAGTGTGACACTTGTGCGGGAGTGTCTATTGTGAAGCTGGCAC 2580  
2672 aagaacccgttatacgaatgaatcacaatagaacaatatacagctatgtcatagcta 2731  
2581 AAGAACCCTGTTACCATGAAATCATCATATTGAAACAATATCAGCTATGTGATTAAGCTA 2640  
2732 tcaagatgaatgactataaagaagaaggtgttgcccaagaacttacaagaagcctt 2791  
2641 TCAAGATGAATGTATCAAGTAAGACAAAGTGTTCACGAACCTTCAAAAGCCCTTT 2700  
2792 cccgttcaagcttcaacttgatgataatgcaatcgtgcccctgtgcaaaagatcctg 2851  
2701 CCCGTTACGGCTTCCACTTGATGATATGSCAATCTGTCCCTTGTGCAAAAGATCCG 2760  
2852 taagagaagaagcctcaatgctgaagcaatgttggtaaaaaataatgaagcgag 2911  
2761 TAAAGAGAGAAAGACCTCATGCTAGGCAATGTTGGTGAATAATATTAAGCGCG 2820  
2912 agtactgaagcagatgagctgttgaatgaataattgtccttccacagagatg 2971  
2821 AGTATCTGAAGCAGATGACGCTGTTAGTAATAATATTGCTCTTCTTACAGAGTATG 2880  
2972 ttgttccatatacaatctcacttctgacacagacagatgltcaaaagtacagata 3031  
2881 TTGTTTCATATACAAATTCACCTTTGGCACATGACCACATATATGCAAGTAAGATA 2940  
3032 ttgaacaacttaagaatgtttaaagaatgtccttggtttcttggaataataatgagcta 3091  
2941 TTGAACAACCTTAAGAATGTAAAGATGCTTTGGTTTGTGGAATAATTAAGGCTA 3000  
3092 aaatgaaataacagtcacgtcttcttacaagaagaatgtgaataatatlaaacaaca 3151  
3001 AAAATGAATAATACAGTCCGCTTTATACGAAGATGTACAAATATTAAACAAACAA 3060  
3152 aagatgcccagaagacagaatgataaagaatgaatgaataaactgtacactgtgtgata 3211  
3061 AAGATGCCCAAGACAGATGATGCAAAATGAATGAATAAATCTGATGATG 3120  
3212 ttgcaatgaatcatcatgtaagaagtaactacatacagtttgaatccttaagacc 3271  
3121 TTGCCATGAAATATCATATGTCMAAGAGTACATACAGTTTGAATCTCTTAAAGACC 3180  
3272 cgttactaacagctgttctcactcaactcaacctggaagaatgtgaatacacaataat 3331  
3181 CGGTACTACACACTGTTTCTTCACTCACTGACAGATTAAGATTTCAGTAACACCAAAAT 3240  
3332 atctgctccttgaataatgaatcatcttcaactcctggaataaactaaacaacaatgttc 3391  
3241 ATCTGCTCTCTGAATAATGAATCATTTTTCATCTCGGAATAACCTTAAACCAATGTTTC 3300  
3392 taggaagctgttaacaagcacttcaacagcagcaagcaatctcaagccaataatcaac 3451  
3301 TAGGAGCTGTTAAACAGCCATTTTCATGACGAGCAAGCAATCTCAGAACCAATATCAC 3360  
3452 gaatgaaacatgtaagaatgcaagcagctgaatccaagcctcctctggaagaataa 3511  
3361 GAATGGAACCTTAAGCAATGCAAGCAGCAGCTCAAAATCCAAAGCTCTCTGGAAGATTA 3420  
3512 aggggaagccttgaatgattctgaatgatacagtggaataatgaatatacacaatgtcct 3571  
3421 AGGGGAGCTTATAGTTCTGAATGATACAGAGTAAGAAATGAAGATTACACAAAGTCTT 3480  
3572 caacttgcggggaanaaagtgacaagaagaagacttgaatcttgaagttgatat 3631



REFERENCE 2 (bases 1 to 5110)  
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-FEB-1999) to the DDBJ/EMBL/Genbank databases. Osamu Ohara, Kazuo DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: odaiinfo@kazuza.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

## FEATURES

source

1. 5110  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="h07056"  
 /clone\_lib="pBluescriptII SK plus"  
 /dev\_stage="adult"  
 /sex="male"  
 /tissue\_type="brain"  
 1. 4253  
 /gene="KIAA0979"  
 <1. 4253  
 /gene="KIAA0979"  
 /product="KIAA0979 protein"  
 /protein\_id="BAA76823.1"  
 /db\_xref="GI:4589602"  
 /translation="VRRLKAVKVTFTMDMDSDSEERELYLNLAHLASDFELKHPDK  
 VRLVACCLADIRIYAPAPYSPDKLIDIFMETIROLKGLDSTSPDRFVLE  
 NIAMWSYNICEFELSDNETFOLYRTLSVINGNHNOKYHMDLMSIICEGTV  
 SOELDPTVNLVPAKRNKOKYDLAKALIKALAEPIYTNFPROVIMLEKTSIS  
 DSEHVFDELLEYNDLSHLVLPOLKELKSNDEEYLOVVKLLAKNEGKDEL  
 ASQNKLMQCYLRNDIHVIRLECVKSHCLMHPDLAKDLTELVKRSDDPEA  
 IRHDVLSIYTAARKDILLVNDHLNFEVERTLDMRWYKEMAGLQYKYALOS  
 AAGDAKQIAWKIKLHLHYONSIDRLVLRVFAOVNPLETEKGLTYLY  
 ATDLNAVKALNEMKCONLRHVOYDLIDIKOPTDAVKAIFSKVWYITNIDNP  
 GKADPMKFTYVLEDEKTRKQLEVVSFTSCQKAGCVREITKLGKOPTNPF  
 LEMIKFLERIAVHIDTESISALIKOVNSIDGTADDEGEVITQDIALRAGLELV  
 LSTFHPISFHSATFESLACLKMDKEVAEALQIFKNTGSIEDFPRIIRALLV  
 LHKSKGPPRAKVAIHCHIAIFSSKEQFAQIFPLHSLDPSNLEHITPLVIG  
 HIALPDQAPALSLVATFIYKIDLMNDRLPGKTKTKLWPDDEVSPTMKIOAI  
 KMWVRLLGKNNKNSGSGTIRLLTTLHSDGLDECKISKPMRSRLAASATV  
 KIAOEPCHYHITLBOYOCALAINDECTIOYROYVQKRLKGLRPLREYATCAL  
 CADPVERKARHROCLVKNINRYRELYLOHAAVSKLSLLEVEVPTLHLLADP  
 DYVKODIDQLDKVCELMFVEILMAKNENSHAFIRKAVENIKQTKDAOGDDAKM  
 NEKLYVCVAMNIIKSKSTYLSLESPKPLVAREFTQPKNSFTKNILPEMKS  
 FPGPKRTNNVGAVPLPSAGKOSOTSSRMETVSNASSNSPSPGKIRKRLDS  
 EMDHSENYETVSSPLPGSKSKOKRSDLDVRSLEKPRGKTKPTVPEOEKILGMDLT  
 KLYOEKPKGSGORSRGRHSHASDEQOMPEERKLEHDLIENDEONSPKKRGR  
 PPLVGCTPREFTMTSKSKGKKKSGPPAPEEEEROSGTEOKSKSHORVSR  
 AQRASPESSAISTOSTPQKGRGRSKTPSPSQPKKNVGRSKQAATKENDSSE  
 VVVFQSSVVDIDPOETEDEEYSTVNVRRRSKRRRR"

BASE COUNT 1714 a 919 c 1045 g 1432 t  
 ORIGIN

Query Match 97.6%; Score 4073; DB 85; Length 5110;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 4076; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 92 tggtagagacattaagaatggttgtaaaactttatgatgatagaccagagcctgaag 151  
 |||||  
 DB 1 TGGTGAACATTAATAAGATGGTGTGAACCTTTATGATGATGACAGACACTCTAAG 60  
 QY 152 aagaagaagagccttattacacctaacttacccttgcctgaagattttttctcaagc 211  
 |||||  
 DB 61 AAGAAAAGAGCTTTATTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTAC 120  
 QY 212 atccctgtaagaatgctcctgctacgctgtaagcctgctgctgctgctgctgctgctg 271  
 |||||  
 DB 121 ATCCTGATAAGATGTCGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
 QY 272 atgctcctgaagcctcctacacacccctgataaactaaagatatattatgcttataa 331  
 |||||  
 DB 181 ATGCTCTGAAGCTCTTACACATCCCTGATAAATAAGGATATATTATGTTATAA 240

QY 332 caagacagttgaagggtctagagatataaagagcccaatcaatgaattatltatc 391  
 |||||  
 DB 241 CAACACAGTTGAAGGGCTAGAGATACAAAGACCCACAGCTTCAATGATGATTTTATT 300  
 QY 392 tacttgaagaatctgctgggtcagaatcatalaacaatccttgagttagaatagca 451  
 |||||  
 DB 301 TACTTGAAGACATGCTTGGGTCAAGTCATATATACATATGCTTGGATTTGAAGATGCA 360  
 QY 452 atgaaatttcacccaactatacagaacctattttcaagtlataaagaagcccaatc 511  
 |||||  
 DB 361 ATGAAATTTTCACCCACTATACAGAACCTTATTTAGTTATAAACAATGGCCCAATC 420  
 QY 512 agaaagtcacatcacaatgtagaaccttatgaacctatatttggaaagtgatcacg 571  
 |||||  
 DB 421 AGAAGTCATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 QY 572 tgcctcagagccttttgatacggctttagtaactcgtgtaacctgctcatalaagaattaa 631  
 |||||  
 DB 481 TGCTCAGAGCTTTTGATACGCTTGTAGTAATCTGTAACCTGCTCATTAAGAAATTTAA 540  
 QY 632 acaagcagcatatgatattggcaagccttactgaagagagacagcccaagctattgagc 691  
 |||||  
 DB 541 ACAAGCAGCATATGATTTGGCAAGGCTTTACTGAAGAGACAGCTCAAGCTATTGAGC 600  
 QY 692 cataattacacacttttttaacatcaggttcctgagcttggaagaacatcatalcagcag 751  
 |||||  
 DB 601 CATATATTACCAATTTTATATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
 QY 752 tgcagaagcagcttctgacttaatttggagcctcacaatattgataagcattgctgc 811  
 |||||  
 DB 661 TGTCAGAGCATGCTTTGACTTAATTTTGGAGCTCTCAATATGATGATGATGATGATG 720  
 QY 812 tctctgttttaaccccaagcttgaatttaataaagagcaatgataatgagagcagctac 871  
 |||||  
 DB 721 TCTCTGTTTAAACCCACTGTAATTTAAATTAAGACCAATGATAAGAGAGGAGGCTTAC 780  
 QY 872 aagttgtaaacctcctgcaaaaatgcttggggaagaagatcagaattgctctcaca 931  
 |||||  
 DB 781 AAGTTGTTAAACTCTGCGCAAAAATGTTTGGGCAAGAGATTCAGATTTGGCTTCTAAA 840  
 QY 932 acaagcactttggcagctgacttggcaggtttaaagatcatalccatccagcc 991  
 |||||  
 DB 841 ACAAGCAGCTTTGGCAGTGTACTTTGGGAGGCTTAAATGATATCCATGATCCGATC 900  
 QY 992 tggagttggaatttggcagcctatgctcctgaacacatcccgatttggcaaaaact 1051  
 |||||  
 DB 901 TGGATGTGTGAATTTGCTAGCCATTTGCTCATGAACCAATCGATTTAGCAAAAAGACT 960  
 QY 1052 taacagagtaacttaagtgagtgacacatgacccctgggaagcatttagacatgatt 1111  
 |||||  
 DB 961 TAACAGAGTATCTTAAGTGTAGTGTACATGACCCCTGAGGAAGCTATTAAGATATG 1020  
 QY 1112 ttgtgtaaatgattacagcctgctaaagaatattccttgcgtcaatgataactactta 1171  
 |||||  
 DB 1021 TTGTGTAATGATTGACAGCTGCTTAAAGATATTTCTTGTGTAATGATCAGTACTTAA 1080  
 QY 1172 atttggagaagagaacatttagaacaagaatggaagatgagcaagaagcagctgctg 1231  
 |||||  
 DB 1081 ATTTTGTGAGAGAGAGACATTAAGCAAAACATGAGAGAGCAAGAGCCATGTATGG 1140  
 QY 1232 gacttcccaaatataagaatattgcttacaagtaagcagcctgggaagaatctgcaa 1291  
 |||||  
 DB 1141 GACTTGGCCCAATTTATTAAGAAATATGCTTTACGTGACGACGCTGGAAGAGATCTGCAA 1200  
 QY 1292 aacagatagcatgatacaagaacaaatgctacatalatalatalcaaaatagattgaty 1351  
 |||||  
 DB 1201 AACGATAGCATGATGATCAAGACAAATTTGCTACATATATATATCAAAATATGATGATG 1260  
 QY 1352 atgactacttgttgaacgagatcttgcctcaatcattggttcccaaatltgaaacta 1411  
 |||||  
 DB 1261 ATGACATCACTTTTGAACGAGATCTTGTCTCAATCATGATGATGATGATGATGATGAT 1320



OY	2341	tttgcgtcctcttggaaccttggtagcgttacttattgtaagaagcttctctatgat	2400
Db	2406	tttgcgtcctcttggaaattcttggtagcttacttattgtaagaagcttctctatgat	2465
OY	2401	gatcgctctccagggaataaagacaactaaacttgggttccagatgaaagatctcct	2460
Db	2466	gatcgctctccagggaataaagacaactaaacttgggttccagatgaaagatctcct	2525
OY	2461	gaagcaatggttcaaaatctcaggtctattaaagatggttcgtagtctacttggaa	2520
Db	2526	gagacaaatggttcaaaatctcaggtctattaaagatggttcgtagtctacttggaa	2585
OY	2521	aataatcacagtaaatccaagaaacttctacctaagaatgcttaacaacaatatgcatagt	2580
Db	2586	aataatcacagtaaatccaagaaacttctacctaagaatgcttaacaacaatatgcatagt	2645
OY	2581	gatggagacttgcacgaacaggggaaattagttaaaccagatgltcaagcttgcagact	2640
Db	2646	gatggagacttgcacgaacaggggaaattagttaaaccagatgltcaagcttgcagact	2705
OY	2641	gctctcggaggtgcttcttggaaagctggagacaagaacctgttaccatgaataatcaaca	2700
Db	2706	gctctcggaggtgcttcttggaaagctggagacaagaacctgttaccatgaataatcaaca	2765
OY	2701	ttagaacaatatcagctctatgttcattagctctatcaacgatagtgctctcaagtaagaca	2760
Db	2766	ttagaacaatatcagctctatgttcattagctctatcaacgatagtgctctcaagtaagaca	2825
OY	2761	gtgttggcccaagaacctccaagaagccttcccggtttagcggcttccacttgatatagt	2820
Db	2826	gtgttggcccaagaacctccaagaagccttcccggtttagcggcttccacttgatatagt	2885
OY	2821	gcaatcctgctccttcttggcaaaagctccctgtaaaagagaagaagcccaatgcagtagca	2880
Db	2886	gcaatcctgctccttcttggcaaaagctccctgtaaaagagaagaagcccaatgcagtagca	2945
OY	2881	tggttggtagaaaataataaattgtaagcgggagatctgaaagcagcatgacgtgttagt	2940
Db	2946	tggttggtagaaaataataaattgtaagcgggagatctgaaagcagcatgacgtgttagt	3005
OY	2941	gaaataattgtgctctcttccacaagagttatgtgttccatatacaaatcttaacttltgca	3000
Db	3006	gaaataattgtgctctcttccacaagagttatgtgttccatatacaaatcttaacttltgca	3065
OY	3001	catgaccacagatctatgtccaagagtagcagatatgtcaacaacttaagaatglttaagaagct	3060
Db	3066	catgaccacagatctatgtccaagagtagcagatatgtcaacaacttaagaatglttaagaagct	3125
OY	3061	cttgggttctgttcttggaatatataatggtctaaaatgtaaataacagtcacgcttctatc	3120
Db	3126	cttgggttctgttcttggaatatataatggtctaaaatgtaaataacagtcacgcttcttattc	3185
OY	3121	agaaagaatgttagaataatatacaacaacaagaatgtgcccgaagccagatgtagtcaaaa	3180
Db	3186	agaaagaatgttagaataatatacaacaacaagaatgtgcccgaagccagatgtagtcaaaa	3245
OY	3181	atgaaatgaaanaactgtacactgtgtgtgtatgttgcacatgaatatatacatgttcaagaagt	3240
Db	3246	atgaaatgaaanaactgtacactgtgtgtgtatgttgcacatgaatatatacatgttcaagaagt	3305
OY	3241	actacataagatttggatctccttaaagcccggtacacacacgcttcggttcttaactca	3300
Db	3306	actacataagatttggatctccttaaagcccggtacacacacgcttcggttcttaactca	3365
OY	3301	ccttacaagaatttccagtlaacacacaaaatactatgcctccctcgtgaatgaaatcatcttcc	3360
Db	3366	ccttacaagaatttccagtlaacacacaaaatactatgcctccctcgtgaatgaaatcatcttcc	3425
OY	3361	actccgtggaaaacctaaacaacaactgttctcagtagagctglttaacaagccacttcatca	3420
Db	3426	actccgtggaaaacctaaacaacaactgttctcagtagagctglttaacaagccacttcatca	3485

OY	3421	gcaggcaaggaacttcacgaccaatcatcacgaatggaaacgttgaagtccaagcgc	3480
Db	3466	GCAGGCAAGCAATTCACAGCCAAATCATCAGCAATGGAACTGTAGCGAATGCAGCACGCC	3545
OY	3481	agctcaaatccaaagctctctcttgaaagaaataaaggaggagcttgatgtctcgaaattgat	3540
Db	3546	AGCTCAAAATCCAAAGCTCTCCTTGGAAGAATAAAGGAGCGCTGTGAATGTTCTGAAATTGCAT	3605
OY	3541	cacagtgaagaatgaagattacaacaatgtcttcacctttgccggggaaaaaagtgcacag	3600
Db	3606	CACACTGAAAATGAAGAGATTACACAATGTCTTCCACTTTGCCGGGAAAAAGAGTGCACAG	3665
OY	3601	aagagcgagctctgtaactcttgtaaagctctaatttgaaagaccataagcgacgaaaanaacg	3660
Db	3666	AGAGACGACTCTGATCTTGTGAAGCTGTGAATTTGGAGAAAGCCTAGAGGCGAGAAAAAACG	3725
OY	3661	cccgtcacagaacagaggagaaattaagtgatlgaaatgacatgaactgaagttggtlacaggaa	3720
Db	3726	CCCCTCACAGAAACGGAGAGAGAAATTAAGTATGTAGTGCCTTGACTTAAGTGTGTACAGGAA	3785
OY	3721	cagaaacctaaaggagagctcagcgaagctcgaaaagaagccataagcgtctcgaactctgat	3780
Db	3786	CAGAACTCTAAAGGACAGTACAGCAAGTGTGGAATAAGGCGCATACGGCTTCAGAAATCTGAT	3845
OY	3781	gaacagcagctggcctcgaaagaaaggagctcacaagaagaatatltagaataatgaatlgaa	3840
Db	3846	GAAAGCAAGTGGCCTTGAGGAAAAAGAGCGCTCAAAGANATATATTGAATAATCAAGATGAA	3905
OY	3841	cagaatagctcgccccaaaaaagggtlaaaggagcggaacccaacacccctctggttgagtc	3900
Db	3906	CAGAAATAGTCGCCCAAAAAAGGTATAAAGAGGCCGACACCAAAACCCTTGTGTGAGAGT	3965
OY	3901	acaccaaaagaagaaccacaatgnaaacctctlaaaaagaagaacaanaaaatcttga	3960
Db	3966	ACACCAAAAGAAAGACCAACATGAAACTTCTTAATAAAGAGACGCAAAAAAATCTGTGA	4025
OY	3961	cctccagcaccagagaggaggaaagaagaagaacaagtgtgaatacggaaacagaag	4020
Db	4026	CCTCCAGACACGAGAGAGAGAMAAACAAACAAAGGTGGAATAATCCGAACACAGAAC	4085
OY	4021	tccaaagaacaaacagcacccgagtgctcaaggagagacagcagagcgagaaatctcttcaa	4080
Db	4086	TCCAAAGCAAAACGACCCGAGTGTCAAGAGAGACGACACGACAGCAGAAATCTTCTGAA	4145
OY	4081	tctagtgcatttgaaatccacacagctccacacacagaagaagacgagagaacatcaaaa	4140
Db	4146	TCTAGTGCAAATGTGAATCCACACAGTCCACACCCACGAAAGACGAGAAACATCATAAA	4205
OY	4141	acggcatcaccatcaccacacacaaaaaatgtg	4173
Db	4206	ACGCCATCACCATCACACCAAAAAAATATGTG	4238
RESULT 3			
LOCUS	ABO23196	5110 bp mRNA	PRI 16-JUN-1999
DEFINITION	Homo sapiens mRNA for KIAA0979 protein, partial cds.		
VERSION	ABO23196		
KEYWORDS	ABO23196.1 GI:4589601		
SOURCE	Homo sapiens adult male brain cDNA to mRNA, clone_1lb;plu:uscriptIIIRI SK plus clone:hj07056.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (sites)		
AUTHORS	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosewa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohnara,O.		
TITLE	XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro		
JOURNAL	DNA Res. 6 (1), 63-70 (1999)		
MEDLINE	99240603		



186 ACTTTTGTGATATGACACAGAGCTGTAAGAAAGAGAGCTTTATTTAAACCTAGCT 245  
181 ttacatcttgcttaagaattttttctcaagatccttgcttaagaattttctgcttactgta 240  
246 TTACATCTTGCTGATGATTTTCTCAAGATCCTGATTAAGATGTTCTGCTTACTGGTA 305  
241 gctgctgcttgcttgatattttcaagattatgctcctgaagctccttacaacacct 300  
306 GCGCTGCGCTTGCTGATTTTCAAGATTTATGCTCTGTAAGCTCCTTACACATCCCT 365  
301 gataactaagaatataatttgatttataaagaagcttgagggtctagaggataca 360  
366 GATTAATTAAGATATATTTATTTATTAACAAGACAGTGAAGGGGCTAGAGATACA 425  
361 aaagagcccaaatcaatgaatttttttacttactagaacatttggttgtaagtca 420  
426 AAGAGCCCAAGCTCAATAGATATTTTATTTGCTTGAAGACATTTGCTGGGTAAGTCA 485  
421 tatacatatgcttgatagaatagaatgaatgaatatttaccacagctatagaagacc 480  
486 TATACATATGCTTTGATTTAGATAGATAGCAATGAATTTTACCCAGCTTACAGAAC 545  
481 ttatttcaagttataaacaatgcccacaatcagaagtcacatgacatgtagacctt 540  
546 TTATTTTCAAGTTATTAACAATGCGCACATCGAAGTCATATGACATGATGACCTT 605  
541 atgagctctatatttttgaaagtgatagaagtgctcagaagcttttgagaaagcttta 600  
606 AAGAGCTCTATTTATTTGTAAGGATGATACAGTGTCTGAGAGCTTTTGATAGGTTTAA 665  
601 gtaaatctgtagacctgctcacaagaattlaaacaagaagaatatttgcaaaagct 660  
666 GTAATCTGATGCTGCTGATTAAGATTTAAACAAGACAGATGATTTGGCAAGGCT 725  
661 ttactgaagaagacagctcaagctatttgagccatatattaccacttttttaacaggtt 720  
726 TTACTGTAAGAGGACAGCTCAAGCTATTGAGCATATTTACCAATTTTAAATCAGGTT 785  
721 ctgagtgcttgagaaacaatcacaagcatttgcagaagcttgcttgacttaatttg 780  
786 CTGATGCTTGAGAAACATCTATCAGCATTTGTCAGAGCATGCTCTGACTTTAATTTTG 845  
781 gagctcacaatatgtagcatttgctgctcctgcttaccacagcttgaattttaa 840  
846 GAGCTCACATATTTGATAGTATGATTTGCTGCTCTGTTTACCAGCTTGAATTTAA 905  
841 ttaagaagcaatgaatgaagagcgcctacaagttgtaaaactatgycgaaaaatgctt 900  
906 TTAAGAGCAATGATTAATGAGAGCGGCTTACAGTTGTTAAACTACGCGCAAAATGTTT 965  
901 ggggcaagaagattcagaattgcttcccaaaaagcgaacttggagtgcttacttgggc 960  
966 GGGGCAAGGATTTAGATTTGCTTCTCAAAACAGCCACTTGTGCACTGCTTACTGGGC 1025  
961 aggttaattatatacctatgtaacaaatccgccttggaattgtaaaatttgctcagccatt 1020  
1026 AGGTTTATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1085  
1021 ctcaatgaacatcctgatttagcaaaaagacttaacagagatctttaaagtgaagctacat 1080  
1086 CTCATGAGCAATCTCTGATTTTGAAGAAAGCTTAAACAGATGATGATGATGATGAT 1145  
1081 gacccatgaagaagcttatagaagatgatttgctcaaatgaattgaagcgtcctaaag 1140  
1146 GACCTTAAGAGAGCTATTTAGATGATGATGATGATGATGATGATGATGATGATGAT 1205  
1141 gatactctctgctcaatgtagcacttaacttaatttgtaagaagagagataatgacaaa 1200  
1206 GATATTTCTTCTGCTCANTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1265  
1201 cgaatgagagtagcgaagaagcaatgtaggagcttgcccaaatltaagaatatgct 1260  
1266 CGATGAGAGTAGACCAAGAAAGCCATGATGGAGCTCCCAAAATTTATTAAGAAATATGCT 1325  
1261 ttacagtcagcagcttggaagaaatgctgctcaaaaagataagcatgtagcacaagaattg 1320  
1326 TTACAGTCAGCAGCTGGAAGAAAGATGCTGCAAAACAGATAGCATGATGATGATGATGAT 1385  
1321 ctacatatataatcaaaaatgattatgtagatcagacttcttgtaagcagatcttgc 1380  
1386 CTACATATATATTTATCAAAATAGATGATGATGATGATGATGATGATGATGATGATGAT 1445  
1381 caatacatggttctctcaaatattagaactcagaagaagatgaatgaatgcttatattctg 1440  
1446 CAATACATGTTGCTCTCAAAATTTAGAACTACAAAGGATGAAATGCTTATATTTACTTG 1505  
1441 tatgcaacactggaatttaaatgctgctgaagaacatttgatgaatgtaattgtaaatg 1500  
1506 TATGCACTGATGATTTAATGCTGTAAGACATTTGATGATGATGATGATGATGATGAT 1565  
1501 ctgctccgacacacaaagaaatgcttctgacttgatttaagcaaccacaaacagatgct 1560  
1566 CTGCTCGACATCAAGTAAGATTTGCTGATTTGATTTAAGCAACCCAAACAGATGCTC 1625  
1561 agtgcaagcacaattttcaaaaatgtagtatttataagaatttaactgctgctgct 1620  
1626 AGTGTCAAGGCTATTTTCAAAAGTGTGTTTATCAAAATTTAACCCTGATCTGCT 1685  
1621 aagctcagaagttcatgaagaataatcaacaggttgattagaagaatgaagaataaga 1680  
1686 AAGCTCAGGATTTTCAATGAGAAATTTACACAGATGTTAGACATGATGAGAAATTAAGA 1745  
1681 aagcagttagaagacttgtagtccaaatgctcctgcaagcagctgaaagcttgctg 1740  
1746 AAGCAGTTAGAAAGTATGTTGATGTCACAAATGCTGTCGCAAGCAGGCTGAAGGTTG 1805  
1741 cgtgaataactaagaagttggtgcaaccccaagaagcctcaaaatcttctctggaatg 1800  
1806 CGTAAATTAACATAAGATTTGGCAACCCCAAAAGCCTTCAAAATCTTCTCTGGAATG 1865  
1801 atcaagttctcttggaagaatagacaccttgacataagaatacgaatcctcaagct 1860  
1866 ATCAAGTTTCTTGTGAGAGGATAGCACTGTGCACATAGTACCGATCTATATAGTCT 1925  
1861 cttaataacaagtgaaacaatcaatagatgaaacaagaatgaatgaatgaatgct 1920  
1926 CTTATTAACAAGATGAAACAAATATGATGATGATGATGATGATGATGATGATGATGAT 1985  
1921 ccaactgatacgaacatcagaagcagcttgaaactgcttgaagctcctcaatcaat 1980  
1986 CCAACTGATCAAGCCATCAGAGAGGCTTGAACCTGTTAAGTATCTCATTTTACACAT 2045  
1981 ccaatcactcaatctctgctgaagaatattgaatcattactgctgcttgaagaatg 2040  
2046 CCATCTCATTTTCAATCTGCTGAAACATTTGAATCATTTACTGCGCTGCTCAAAATGAT 2105  
2041 gatgaagaagtagcagaagctgcaatcaaaatttcaaaaacacagaagaagaattgaa 2100  
2106 GATGAAGAAAGTAGCAGAAAGCTGCACTACAAATTTTCAAAACACAGGAAGCAAAATGAA 2165  
2101 gagatcttccacacatcaagctgctgctgctgctgctgcttcaatcaaaatcaaaa 2160  
2166 GAGATTTTCCACACATCAGATCAGGCTTGTCTCTGTTTATCATCAAAATCTTAAAAA 2225  
2161 ggaaccccccgtcaagaacaaatgcaatcaatgtaattgcaatgcaatgcaatgcaat 2220  
2226 GGACCCCGCTCAAGCAAGCAATATGCTTATTTGATGATGATGATGATGATGATGATGAT 2285  
2221 gagaacccagtttgcaagataattgagcctctgtaagaagcctagatcccaagaacctg 2280  
2286 GAGGCCAGTTTGGACAGATATTTGAGGCTCTGATTAAGACCTAGATCCAAAGCAACCTG 2345  
2281 gaacatctcataacaacatggttactatttggtcatatattgctcctctgcaactgata 2340  
2346 GAACATCTCATTAACACATTTGTTACTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2405



CDS

/number=2  
66..4241  
/gene="AS3"  
/codon\_start=1  
/product="androgen-induced prostate proliferative shutoff  
associated protein"  
/protein\_id="AND22134.2"  
/db\_xref="GI:4559410"  
/translation="MAHSKTRTNDGKITPPGVKEISDKISKENAVRLKMKVKTMD  
MDQSEEEKELYLNALHLASDFLKHDPKDVRLVACCLADIFRIYAPVATSPDK  
LKIDFMFTIRLOLKGLEDKRSQPNFNYFYLENLIAWKGYNICFELSDNETIETOLYRT  
LESVYNNGNHOKVHHMWDMSITICEEDTYSOELLDVYVNLVPAHKNLKOATDLA  
KALKRTAOATEPITITNFNOVLMGKTSISDSHVSDDLIELYNIDSHLLSVLPQ  
LEFKLSNDNEERLAQVRLAKMFGAKDSELAQKPLMQCYLGRFNDIHPRIELCY  
VFASHCLNNHDPDLADLTEYLKVRSHDEEALRHVIVISYTAARKDILLVNDILEY  
VGEERTLDKRMVRYKEMAMGLAOIYKKVALQSAAGDAKQAIAMIKDKLHLHYOSID  
DLIVERLIFAQYVPMHNLETTERMKCLYIYATLDLNAVKALNEMWKKONLIRYOSID  
LLDLTKPKRTDASVKAIFSKVYVITRSLPDRKADPPMKKTQVLEDEKIRKQLEVL  
VSPSCQADBCVREITKLGKPKOPNPLEMKTELLERLAPVHIDESTISALIKO  
VNSISDGTADDEGVPTDOAIRAGLELLKVLSETHPISFHSAEFESILACLKMDDE  
KVAEALQIFPKNTGSKIEDEPHIRSALEPVLHHSKRGPPROAKYAIHCIIAIFSSK  
EAOFAOIEPRLKSLDPNSNLEHLITPLVTIGIALADPOFAAPLKSIVATFVLDL  
MNDRLPGKTKTKIWPDEEVSPEPMVKTOAIKMWVRMLLCKMNHSSKSGTSLRLTT  
ILHSDGLTEBOKISKPDMSRLRLAAGSAIVKLAQPCYHEITILEQYQCALAINDE  
CYQVQVFAOKLHKLSRLRPLEYMAICACAKDVERRAHARQCLVKNINVREY  
LKQHAASSEKLSLEPEYVPTIHLAHADPYVODIEOLKDYKELMPELITMA  
KNENSHAFIRKMNENIKOTKDAQPDDAKMKNEKYTCDAVAMNIIMSKSTYSLESP  
KDPVLPAREFTQDPKFNSTNKNYLPKPGKSFETPGKPKTNTVLGAVNKLPSAGKQSO  
TKSSRMETVGNASSSNPSPGRIKGRDSSPMHSEMDYTMSPLPGKSRSDKDDSD  
DLYRELEKPRGRKTPYTEREKEKGMIDLTKLVQEQRPKGSQSRKRGHINASESDQ  
QMPERKRLKEDILENEDQNSPPKGGKGRPKPLGGSTPEEPTMKSKSGSKRSKSG  
PPAPEEEEEESGNTQKSKSKOHVRSRAQNAESSESATISTOSTQKGRNP  
SKTSPSQPKKNV"  
174..377  
/gene="AS3"  
/number=3  
378..464  
/gene="AS3"  
/number=4  
465..562  
/gene="AS3"  
/number=5  
563..689  
/gene="AS3"  
/number=6  
690..770  
/gene="AS3"  
/number=7  
771..911  
/gene="AS3"  
/number=8  
912..1027  
/gene="AS3"  
/number=9  
1028..1122  
/gene="AS3"  
/number=10  
1123..1268  
/gene="AS3"  
/number=11  
1269..1420  
/gene="AS3"  
/number=12  
1421..1534  
/gene="AS3"  
/number=13  
1535..1616  
/gene="AS3"  
/number=14  
1617..1665  
/gene="AS3"  
/number=15  
1666..1805

/gene="AS3"  
/number=16  
1806..1921  
/gene="AS3"  
/number=17  
1922..2027  
/gene="AS3"  
/number=18  
2028..2188  
/gene="AS3"  
/number=19  
2189..2312  
/gene="AS3"  
/number=20  
2313..2471  
/gene="AS3"  
/number=21  
2472..2540  
/gene="AS3"  
/number=22  
2541..2677  
/gene="AS3"  
/number=23  
2678..2801  
/gene="AS3"  
/number=24  
2802..3006  
/gene="AS3"  
/number=25  
3007..3121  
/gene="AS3"  
/number=26  
3122..3254  
/gene="AS3"  
/number=27  
3255..3374  
/gene="AS3"  
/number=28  
3375..3437  
/gene="AS3"  
/number=29  
3438..3583  
/gene="AS3"  
/number=30  
3584..3689  
/gene="AS3"  
/number=31  
3690..4129  
/gene="AS3"  
/number=32  
4130..4354  
/gene="AS3"  
/number=33  
4355..4810  
/gene="AS3"  
/number=34

BASE COUNT 1755 a 944 c 1074 g 1480 t  
ORIGIN

Query Match 99.2% Score 4141; DB 97; Length 5253;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 4153; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 atggctcattcaaaacttagagccaatgatagaaaattactatccgctggggtcaag 60  
|||||  
DB 66 ATGGCTCATTCAAAAGACTAGACCAATGATGGAATAATTACTATCCCGCTGGGTCAAG 125  
|||||

QY 61 gaatatcagataaaatctaaagagagatggtgagacgattaaagattgtgaa 120  
|||||  
DB 126 GAAATATCAGATAAATATCTAAAGAGAGATGTCGACGATTAAGATGGTTGTGAAA 185  
|||||

QY 121 accttatgatalgaccagagactctgaagaagaagaagagcttatttaaacctagct 180



Db 2947 GCATCTGTCCTTGTGCAAAAAGATCTGTAAGAGAGAAAGAGCTCATGTCAGCAAA 3006  
 2881 ttttttgtaaaaaataataatgttaagcggagatcttgaagcgcctgcaagcgttagt 2940  
 Db 3007 TGTGTGTGAAAAATATATATATATAGCGGAGATCTGTAAGAGAGAGCTCATGTCAGCAAA 3066  
 2941 gaaaaatattgtctctcttaccagagatgtgttccataatacattacccttttgca 3000  
 Db 3067 GAAAAATATTATGTCCTTCAACGAGATGTTCTTCATATACAAATTCACCTTTTGCA 3126  
 3001 catgaccagattatgtcaaaagtaacagatatgtacaacttaagaatgttaagaatgt 3060  
 Db 3127 CATGACCCAGATATGTCAAAAGTACAGATATGAAACAACCTTAAAGATGTTAAAGATGT 3186  
 3061 ctttggtttgttctgtaaatattatgtcttaaaaaatgaataacagtcacgtttatc 3120  
 Db 3187 CTTTGGTTTGTCTGGAATATATATATGCTAAATAATGAAATTAACAGTCACGCTTTTATC 3246  
 3121 agaaagatgtgtaagaataatataaacaacaagaatgccagagaccagatgtgcaaaa 3180  
 Db 3247 AGAAAGATGTGAGAAATATTTAAACAAACAAAGATGCCAGAGACAGATGATGCAAAA 3306  
 3181 atgaatgaaaaaactgtacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3240  
 Db 3307 ATGAATGAAAAAATGTACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3366  
 3241 actacatacagttgtggaatctccataaagaccggtactaccagctgttttcttacc 3300  
 Db 3367 ACTATCAATACAGTTTGGATCTCCCTAAAGACCCGGTACTACAGCTCGTTTCTTCACTCA 3426  
 3301 cctacagaagaattcagtaacacacaaatattatctgcctccctgtaaatgaatacttttc 3360  
 Db 3427 CCTACACAGAAATTTCACTACACCAAAATATATCTGCTCTGTAATGAAATTCATTTTTC 3486  
 3361 actctgtgaaaaacctaacaacacacatgttcttagagctgtttaaagccacttcalca 3420  
 Db 3487 ACTCTGGAAGAACTTAACACCAATGTTCTAGAGCTGTTAAACAGCCCTTTTCATCA 3546  
 3421 gcagagcagaatctcagacccaatctcaccgaatgtgaactgttaagcaatgcaagcgc 3480  
 Db 3547 GCAGGCAAGCAATCTGACACCAATCATCAGAAATGGAATGGAAGCAAGCAAGC 3606  
 3481 agccaatccaagctctccctggaagataaaggagcgttgaagttcttgaagatgat 3540  
 Db 3607 AGCTCAATCCAACTCTCTCGGAGAAATTAAGGAGGAGCCTTGATAGTTCTGAAATGAT 3666  
 3541 caccgtgaaaaatgaagattacacacatgtctcacccttgcgagggaataaagtgcacag 3600  
 Db 3667 CACAGTGAATAATGAAGATTACACAAATGCTTCACCTTGGCGGGAATAAAGTGCACAG 3726  
 3601 aggaagacacttgatcttctgttaagttcgaattgtgaagagccttaagcagaagaaaaaag 3660  
 Db 3727 AGAAGACACTCTGTGATCTGTGAAGGCTGGAATTTGAGAGGCTTACAGGAGGAAAAAAGC 3786  
 3661 ccgcctcagaagagaggaagaaattaggtatgtgactgacactgaagtgtgttgcagaa 3720  
 Db 3787 CCCTTCACAGACAGAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATGAT 3846  
 3721 cagaacacttaaggcactcagcgaagtcggaagagcgaacacacacacacacacacacacac 3780  
 Db 3847 CAGAAACCTTAAGGCACTACAGCAAGTCGCAAAAGAGGCCCTACGGCTTCAGAAATCTGAT 3906  
 3781 gaaagcagtggtccttgagaaagagcctcaagaaagatatatagaaatgtaagagaa 3840  
 Db 3907 GAAAGAGAGTGCTGTGAGGAAAGAGGCTCAAGAGATATATATGAAATGAAGATGAA 3966  
 3841 cagaatgtccgcgaacaaagaggttaaaagagcgcaacacacacacacacacacacacacacac 3900  
 Db 3967 CAGAAATAGTCCGCAAAAAAGGTTAAAAAGAGCGGACCAACCAACCTCTTGCTGAGAGT 4026  
 3901 acacccaaagagagcacaatgtaaaacttcaaaaaaggaacaaaaaaactctgga 3960

Db 4027 ACACCAAGAGAGAGCCACATGAAACTCTTAAAAAGCAACCAAAAAATCTGGA 4086  
 3961 cctccagcaccagagagagagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 4020  
 Db 4087 CTTCCAGCACAG 4146  
 4021 tccaaagcaaacagacagcagagtgtaagagagcagacagagagagagagagatctctgaa 4080  
 Db 4147 TCCAAAGCAACAGACACCGAGTGTCAAGAGAGACACAGCAGAGAGAGAGATCTCTGAA 4206  
 4081 tctagtcaattgaatccacacagtcacacacacacagaaagagagagagagagagagagag 4140  
 Db 4207 TCTAGTCAATTTGAATCCACAGCTCCACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4266  
 4141 agccatcacatcac 4172  
 Db 4267 AGCCATCCCATCCATCCACCAAAAAAATGT 4298  
 RESULT 2  
 LOCUS HSU95825 5253 bp mRNA PRI 05-APR-1999  
 DEFINITION Human androgen-induced prostate proliferative shutoff associated protein (As3) mRNA, complete cds.  
 ACCESSION U95825  
 VERSION U95825.2 GI:4559409  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 5253)  
 AUTHORS Geck, P., Szelel, J., Jimenez, J., Soto, A.M., and Sonnenschein, C.  
 JOURNAL Androgen-induced proliferative shutoff in prostate cancer cells  
 TITLE 2 (bases 1 to 5253)  
 REFERENCE 2 (bases 1 to 5253)  
 AUTHORS Geck, P., Szelel, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.  
 JOURNAL Early gene expression during androgen-induced inhibition of proliferation of prostate cancer cells: a new suppressor candidate on chromosome 13, in the BRCA2-Rb1 locus  
 TITLE 3 (bases 1 to 5253)  
 REFERENCE 3 (bases 1 to 5253)  
 AUTHORS Geck, P., Szelel, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.  
 JOURNAL Direct Submission  
 TITLE Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA  
 REFERENCE 4 (bases 1 to 5253)  
 AUTHORS Geck, P., Szelel, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.  
 JOURNAL Direct Submission  
 TITLE Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA  
 REMARK Sequence update by submitter  
 COMMENT On Apr 5, 1999 this sequence version replaced gi:4539617.  
 FEATURES  
 location/Qualifiers  
 1..5253  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /map="13q12-q13"  
 /cell\_line="LNCaP"  
 /cell\_type="cancer cells"  
 /tissue\_type="prostate"  
 /note="exon 1 maps to P1 Artificial Chromosome PAC26H23; exons 2-24 map to cosmid 267p19; exons 25-34 map to P1 Artificial Chromosome PAC 49J10"  
 1..5253  
 /gene="AS3"  
 1..46  
 /gene="AS3"  
 /number=1  
 47..173  
 /gene="AS3"  
 exon



QY	661	ttactgaagggaaagctcaagctatttggccatatataccaccttlttbaatcaagtt	720
Db	787	TTACTGAAGAGAGACAGCTCAAGCTATTGGACCCTAATATTTCCAAATTTTTTAAATACGGTT	846
QY	721	ctgaatgcttggaaaacatcatcaacagcgatttgcagacatgccttgaactaaatttg	780
Db	847	CTGATGCGTTGGGAAAACATCTATCACAGATTTGTCCAGACATGCTTTGACTTATTTTG	906
QY	781	gaagcttacaatatltgaatgaatctgtgcgtcctcgttlttaacccagcttgaattaa	840
Db	907	GAGCTCTACATATTTGATTCATTTGGCTGCTGCTCTGTTTACCCCACTTGAATTTAA	966
QY	841	ttaaagagcaatgtataatgtgagagcgccctacaagattgttaaactctctgccaataatgttt	900
Db	967	TTAAAGAGCAATGTATATGTAGAGAGCCCTTACCAAGTTCCTTAAACTCTGCAAAAATGTTT	1026
QY	901	ggggcaaaagattcaagaatttggctcttccaaaacagccaacttggcaatgtactattgac	960
Db	1027	GGGGCAAAAGGATTCAGAAATTTGGCTTCCAAAACAAACCCATTGGCAGTGCATCTTGGGC	1086
QY	961	aggttctaagatcatcaatgtcaacatccgcctggaatgttgaatatltgttagccattgt	1020
Db	1087	AGGTTTATGTATGTATCCATGTACCAATCCGCTCGEATGTGTCAAAATTTGCTAGCATTTGT	1146
QY	1021	ctcatgaacctccctgttttagaacaagaagcttaacaggaatcttctaagaatgaatgcacat	1080
Db	1147	CTCATGAACCATCCCTGTATTTAGCAAAAAGCTTAAACAGATATCTTAAAGTGAGTCACAT	1206
QY	1081	gaaccttgaagaaactatagaacatgtatgttatgtgtcaatgtttacagcttgcataaag	1140
Db	1207	GACCTGAGGAGAGCTATTACACATGATGTTATGTGTCAATATGTTACAGCTCTCTAAAG	1266
QY	1141	gatatctctctgttcaatgtcatgtcaacttaatttcttgaagagagaacatlagacaat	1200
Db	1267	GATATTTCTTCTGGCATGTATGTACCTTACTTAAATTTGTGAGAGAGAACTATAGCAAA	1326
QY	1201	cgatgagagatlagcaaaagaagcactgatatggagacttggccaaattataagaataatgct	1260
Db	1327	CGATGGAGATGACGCANAGNAGCCATGATGGAGCTTGCCCAATTTATAGAAATATGCT	1386
QY	1261	ttacagctcagcagcttggaaaaagaatgcctgcacaaacagatagcatgtatcaaaagacaattg	1320
Db	1387	TTACAGTCAGAGAGCTGGAAAAGATGCTGCAAAAACATATGATGATCAAGACAAATTG	1446
QY	1321	ctacatatataatatacaaatagatlatatgatatgcgactacttgttgaacgagacttgcct	1380
Db	1447	CTACATTTATATTTATTCMAAATAGTATTTGATGATCGCACTCTCTGTTGAACGGATCTTTTGCT	1506
QY	1381	caatcacatgttctccatcaaatlttagaacttacaagacaggaatgaatgtcttatatactg	1440
Db	1507	CAATACATGGTTCTCTCACAATTTTANAACTACAGAACGATTAATGCTTTATATTACTTG	1566
QY	1441	tatgcacacatggaatttaaaatgtctgtgaaagaacatltgaatgaatgttggaaaatgttcaaat	1500
Db	1567	TATGCCACACCTGGATTTAAATGCTGTGAAAAGCATTTGAATGAATGTGMAATGTCAAAAT	1626
QY	1501	ctgtctcgcacatcaaatgaatgaatgttgccttgcactgataagcaacccaacagatgac	1560
Db	1627	CTGTCTCGCAATCAAGTAAAGGATTTGCTTTGATTTGATTTAACCAACCCAAAACAGATGCC	1686
QY	1561	agttcaaaagccatatttccaagaatgtaatgtatatacaagaatattccgtatccctgct	1620
Db	1687	AGTGTCAAGGCCATATTTTCAAAAAGTATGATGTTATTACANAGAAATTTTACCTGATCTGCT	1746
QY	1621	aagcttcaagatttcatbgaagaaatctacacaggtgttgaagaatgatatgaatgaatgaatga	1680
Db	1747	AAGGCTCAGATTTTCATGAGAAATTCACACAGAGTGTATGAAAGATGATGAGAAATAAGA	1806
QY	1661	aagaagttagaatgaatcttgttgaatccaacatgtcctctcgaagcagagcttgaagttgtgtg	1740
Db	1807	AAGAGTTTAAAGATGACTTGTATGTCCAAATGCTCTCTGCAACAGGCTGAAGGTTGTGTG	1866

QY	1741	cgtagaataactaagaagttggggaacccccaacacccacaacatcccttccttctgtaag	1800
Db	1867	cgtagaataactaagaagttggggaacccccaacacccccaacacatcccttcctgtaag	1926
QY	1801	atcaagttccctctggagagatagaccctgtgacatatagataccgaatctacagtgct	1860
Db	1927	atcaagtttcttcttgagagagatagaccctgtgacatatagataccgaatctacagtgct	1986
QY	1861	cttatcaacaagtgaaccaaataataatagaaagcgcagaatgataagagatgagaggtgtc	1920
Db	1987	cttatcaacaagaaggaaccaaataataatagaaagcgcagaatgataagagatgagaggtgtt	2046
QY	1921	ccaactgtgacaagcatcaagacgaagtccttgaacgtcttaagtaactctcatcttaacat	1980
Db	2047	ccaactgtatcaagccatcaagacgaagtccttgaacgtcttaagtaactctcatcttaacat	2106
QY	1981	cccatctcatctcaatctctgtctgaaacattgaaatcaatctacgtgtgtctgtaaatgtat	2040
Db	2107	cccatctcatctcaatctctgtctgaaacattgaaatcaatctacgtgtgtctgtaaatgtat	2166
QY	2041	gatgaanaaagtagagaagaagctgtgacacaaatttcaaaaacacaggaagcaaatgtga	2100
Db	2167	gatgaanaaagtagagaagaagctgtgacacaaatttcaaaaacacaggaagcaaatgtga	2266
QY	2101	gaggaatttccacacatcagatcagcctgtgtcctgttltacatcacaaatctaaanaa	2160
Db	2227	gaggaatttccacacatcagatcagcctgtgtcctgttltacatcacaaatctaaanaa	2286
QY	2161	ggaaccccccgtagaaccaaatatgcacatctatgtatcccatgagatattttctagttaa	2220
Db	2287	ggaaccccccgtagaaccaaatatgcacatctatgtatcccatgagatattttctagttaa	2346
QY	2221	gagaccgaatttgacacagatatattgagccctgcataaagaagcctgataccaaagcaactg	2280
Db	2347	gagaccgaatttgacacagatatattgagccctgcataaagaagcctgataccaaagcaactg	2406
QY	2281	gaacatctcaataacacatigtgttaactatgtgtcatattgtctctcctgtgacactgata	2340
Db	2407	gaacatctcaataacacacatigtgttaactatgtgtcatattgtctctcctgtgacactgata	2466
QY	2341	tttgctgtcctctggaaactctggagagcgaacttaattgtgaaagaatctctcaatgtat	2400
Db	2467	tttgctgtcctctggaaactctggagagcgaacttaattgtgaaagaatctctcaatgtat	2526
QY	2401	gatcggtctccagggaaaaagacaaactaaacttttgggtctccagatgaaagatgactcct	2460
Db	2527	gatcggtctccagggaaaaagacaaactaaacttttgggtctccagatgaaagatgactcct	2586
QY	2461	gaggaacaatgycaaaatctcaggtctatataatgatgtgtcgatggtcctcttggaaatga	2520
Db	2587	gaggaacaatgycaaaatctcaggtctatataatgatgtgtcgatggtcctcttggaaatga	2646
QY	2521	aataatcaagaataatcaagaactctcacttaagaatgtgtaacaacaatatgtcatgt	2580
Db	2647	aataatcaagaataatcaagaactctcacttaagaatgtgtaacaacaatatgtcatgt	2706
QY	2561	gatggagaacttgacagaaacaggggaaatatagtaaacagatatgtlcaagctctgagact	2640
Db	2707	gatggagaacttgacagaaacaggggaaatatagtaaacagatatgtlcaagctctgagact	2766
QY	2641	gctctctggagtgatatgttgaagcctggcacaagaacctgttaccatgaaatacatca	2700
Db	2767	gctctctggagtgatatgttgaagcctggcacaagaacctgttaccatgaaatacatca	2826
QY	2701	ttagaacaatatcagctaigtgatagcatatcaacagatgaaatgtcatcaagtaagaa	2760
Db	2827	ttagaacaatatcagctaigtgatagcatatcaacagatgaaatgtcatcaagtaagaa	2886
QY	2761	gtgtgttcccgagaactctgcaaaagcctttccogtttcaagccttcaactgtgagatat	2820
Db	2887	gtgtgttcccgagaactctgcaaaagcctttccogtttcaagccttcaactgtgagatat	2946
QY	2821	gcaacttgtgccccttctgtcaaaaagatctcgttlaaagagagaagagccatgctcagaa	2880



exon /number=20  
2374..2532  
/number=21  
2533..2601  
exon /number=22  
2602..2738  
exon /number=23  
2739..2862  
exon /number=24  
2863..3067  
misc\_feature /number=25  
/complement(2912..3302)  
/note="matches EST AA630616 from clone 856164"  
exon /number=26  
3068..3182  
/number=27  
3183..3315  
exon /number=27  
3302..3572  
misc\_feature /note="matches EST AA333118"  
3316..3435  
exon /number=28  
3349..3818  
misc\_feature /note="matches EST AA282190 from clone IMAGE:704631"  
3436..3498  
exon /number=29  
3499..3644  
/number=30  
/complement(3586..4065)  
/note="matches EST AI655429 from clone IMAGE:2239331"  
3645..3750  
exon /number=31  
3751..4190  
/number=32  
/complement(3864..4286)  
/note="matches EST AI680124 from clone IMAGE:2264521"  
4001..4119  
misc\_feature /note="matches EST N83765 from clone KK2745"  
4135..4450  
misc\_feature /note="matches EST F12354 from clone c-38h11."  
4191..4298  
exon /number=33  
4255..4307  
misc\_feature /note="matches EST T64867 from clone 22264"  
4299..4434  
exon /number=34  
/join(4398..4533,4518..4660,4659..4675,4674..4693)  
/note="matches EST H44781 from clone 188032"  
4435..7444  
exon /number=35  
4537..5022  
misc\_feature /note="matches EST AI681422 from clone IMAGE:2272632"  
4537..4892  
misc\_feature /note="matches EST AA911125 from clone IMAGE:1416862"  
4538..4969  
misc\_feature /note="matches EST AI275866 from clone IMAGE:1877482"  
4538..4947  
misc\_feature /note="matches EST AA987361 from clone IMAGE:160350"  
4538..4943  
misc\_feature /note="matches EST AI299646 from clone IMAGE:1898010"  
4362..5003  
misc\_feature /note="matches EST AA447433 from clone 784555"  
4571..4854  
misc\_feature /note="matches EST H93424 from clone 220700"  
/join(4573..4649,4650..4815)  
/note="matches EST W20070 from clone 306049"  
/complement(4529..4937)  
/note="matches EST AW089775 from clone IMAGE:2593146"  
4746..4994  
misc\_feature /note="matches EST AA747568 from clone IMAGE:1269696"  
/complement(4803..5332)  
/note="matches EST AI911784 from clone IMAGE:2329873"  
/complement(4828..5296)  
/note="matches EST AA563884 from clone IMAGE:1013990"

misc\_feature 4861..5243  
/note="matches EST D20082 from clone pm1283."  
misc\_feature /complement(join(4872..4921,5281..5332))  
/note="matches EST AW212954 from clone IMAGE:2647801"  
misc\_feature join(4872..4921,5281..5332)  
/note="matches EST AI177347 from clone 621810"  
misc\_feature /complement(4872..4921)  
/note="matches EST AW048929 from clone  
UI-M-BH1-amo-c-09-0-UI"  
4872..4921  
misc\_feature /note="matches EST AA154932 from clone 599194"  
/complement(4876..5332)  
misc\_feature /note="matches EST AI016896 from clone IMAGE:1627886"  
/complement(join(4922..4975,5281..5332))  
/note="matches EST AI012649 from clone RPLBA02"  
misc\_feature /complement(4922..4975)  
/note="matches EST AA899586 from clone  
UI-R-E0-cy-e-02-0-UI: matches EST AI172172 from clone  
RM0BV82"  
misc\_feature /complement(4931..5332)  
/note="matches EST AA625961 from clone 745478"  
misc\_feature /complement(4951..5332)  
/note="matches EST AI140790 from clone IMAGE:1566222"  
misc\_feature /complement(5021..5309)  
/note="matches EST H45693 from clone 188292"  
  
Query Match 99.8%; Score 4165.6; DB 92; Length 7444;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 4168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 atggctcattcaaaagactaagaccatgatgaaataatcatatccgcctgaggtcaag 60  
|||||  
DB 127 ATGGCTCATTTCAAGACTAGACCAATGATGAAATAATTCATATCCGCTGGGCTCAAG 186  
  
QY 61 gaaatcaagaataaataatcctaaagaagagatggttgagcagataaagaatggttgaaa 120  
|||||  
DB 187 GAAATATCAGATTAATAATATCTAAAGAGAGATGCTGAGACATTAAAGATGCTTGTAAAA 246  
  
QY 121 actttatgtagatgagccagactctgaaagaataaagaagacttttaaacctagct 180  
|||||  
DB 247 ACTTTATGATATGAGCCAGACTCTGAAAGAAAAGAGCTTTATTTAAACCTAGCT 306  
  
QY 181 ttacatctgctcagatlttttctcaagacatcctgtaagaatggtcgtctactgta 240  
|||||  
DB 307 TTACATCTGCTTCAGATTTTCTCAGACATCTGATTAAGATGTTCCGTTACTGTA 366  
  
QY 241 gectgctgctctgctgataatctcaagaatlaagctcctgaagctccttacacatccct 300  
|||||  
DB 367 GCCTGCTGCTTGTCTGATTTTTCAGGATTTATCTCCTGAAGCTCCTTACACATCCCT 426  
  
QY 301 gataactaaagatatattatlttatacaagacagttgaaggagctagaaggtatata 360  
|||||  
DB 427 GATTAAGCTTAAGGATATATTTATGTTTATATACAGACAGTTGAAGGGGCTAGAGGATACA 486  
  
QY 361 aagagccacaatcaatagaglatlttatttacttgaacaatgcttgggtcaagltca 420  
|||||  
DB 487 AAGAGCCCAATTCATATAGTATTTTATTTACTTGGAACATTTGCTGGGCTCAAGTCA 546  
  
QY 421 tataacataagcttggatgaagaatagaatgaatattccaccagctataagaacc 480  
|||||  
DB 547 TATAACATATGCTTGTGATTTGAAGATAGCAATGAATTTTACCCAGCTTATACAGAAC 606  
  
QY 481 ttatttcagttataaacaatgagccacaatagaagaatgcataatgacatgtagacct 540  
|||||  
DB 607 TTATTTTCAGTTATTAACATAGCCACATACAGAAAGTCCATATGACATGTGTAGACTT 666  
  
QY 541 atgagctctatatttggaagtgatacagtgctcagaagcttggatgaaggttca 600  
|||||  
DB 667 ATGAGCTCTATTTATTTTGGAAGTGATACACTGTCTGAGAGCTTTTGATGAGGTTTAA 726  
  
QY 601 gtaaatctgtagcctgctcacaagaatttaacaagaagcatatgattggcaaggtc 660  
|||||  
DB 727 GTAATCTGTACCTGCTCATTAAGATTTTAAAGCAAGCATATGATTTGGCAAGGCT 786



C 9	443	10.6	72157	64	AC016449	AC016449 Homo sapi
C 10	389.8	9.3	54398	73	AC068224	AC068224 Homo sapi
C 11	232.2	5.6	2212	89	AK021757	AK021757 Homo sapi
C 12	207.4	5.0	54398	73	AC068224	AC068224 Homo sapi
C 13	206.8	5.0	113704	92	HS26P19	275889 Human DNA s
C 14	206.8	5.0	163642	79	AL353724	AL353724 Homo sapi
C 15	179.6	4.3	193880	80	AL358892	AL358892 Homo sapi
C 16	179.6	4.3	340969	81	AL512630	AL512630 Mus muscu
C 17	163	3.9	163642	79	AL353724	AL353724 Homo sapi
C 18	131	3.1	145450	73	AC068352	AC068352 Homo sapi
C 19	131	3.1	160962	67	AC022862	AC022862 Homo sapi
C 20	131	3.1	167621	77	AC090081	AC090081 Homo sapi
C 21	124	3.0	69740	77	AC087748	AC087748 Homo sapi
C 22	121.4	2.9	145450	73	AC068352	AC068352 Homo sapi
C 23	121.4	2.9	162775	67	AC023006	AC023006 Homo sapi
C 24	119.2	2.9	173613	4	AC007475	AC007475 Drosophill
C 25	119.2	2.9	192763	4	AC007474	AC007474 Drosophill
C 26	119.2	2.9	194634	65	AC020286	AC020286 Drosophill
C 27	119.2	2.9	262731	5	AE003823	AE003823 Drosophill
C 28	115.4	2.8	69740	77	AC087748	AC087748 Homo sapi
C 29	108	2.6	1489	89	AK023592	AK023592 Homo sapi
C 30	106	2.5	72157	64	AC016449	AC016449 Homo sapi
C 31	99.8	2.4	7218	10	166494	Sequence 14
C 32	98.6	2.4	162775	67	AC023006	AC023006 Homo sapi
C 33	87.6	2.1	167671	77	AC090081	AC090081 Homo sapi
C 34	83.8	2.0	160962	67	AC022862	AC022862 Homo sapi
C 35	70.8	1.7	978	53	CNS06PJK	AL409446 T7 end of
C 36	68.6	1.6	156550	64	AC015830	AC015830 Homo sapi
C 37	68	1.6	147146	83	AP003316	AP003316 Oryza sat
C 38	67.6	1.6	66288	78	AC090825	AC090825 Homo sapi
C 39	66.4	1.6	886	53	CNS075DF	AL429961 Clome BA0
C 40	65.6	1.6	153665	78	AL138818	AL138818 Homo sapi
C 41	65.4	1.6	23502	6	CEL2K355	AC006832 Caenorhab
C 42	65.2	1.6	74119	71	AC036177	AC036177 Homo sapi
C 43	64.8	1.6	42839	70	AC027282	AC027282 Homo sapi
C 44	64.6	1.5	68844	70	AC027155	AC027155 Mus muscu
C 45	64.6	1.5	81120	67	AC022851	AC022851 Homo sapi

## ALIGNMENTS

RESULT 1  
 HS26H231 7444 bp mRNA PRI 25-JAN-2000  
 DEFINITION Novel human gene mapping to chromosome 13.  
 ACCESSION AL137201  
 VERSION AL137201.1 GI:6759511

## SOURCE

ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 7444)  
 TITLE Rhodes, S. and Huckle, E.  
 JOURNAL Direct Submission  
 COMMENT Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk  
 This cDNA sequence was assembled from public domain ESTs and single  
 pass sequencing reads from expressed DNA templates, aligned to the  
 genomic DNA sequence from the bacterial clones 26H23 (Z84467),  
 267P19 (Z75889) and 49J10 (Z84572).

The EST sequences listed match this sequence with an identity of at  
 least 95% between the coordinates shown.  
 Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene  
 Sanger Centre name: 26H23 Cl3.1.

## FEATURES

source  
 1..7444  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /map="13q12.3"  
 1..107  
 exon

exon	/number=1 108..234 /number=2 127..4470 /codon_start=1 /product="hypothetical protein" /protein_id="CAB69911.1" /db_xref="GI:6759512"
CDS	translation="MAHSTRNDGKITPPGVKEISDKISKEEMVRRLKMKVTFMD MDQSEEEKLEYLNALASDFELKHPDKOVRLLVACCLADIPFIVAPAPYSPDK LKDIEMFTROLKGLDEPKSPQPNFYFLLENIMAVSYNCFLELSDLEIETQYRT LFVYINNGHNNQVHHHNDLMSSTICEEDTYSQELDPVLYNLVPAHNLNKQATDIA KALKRTAOALEPYITNFFNOYLMIGKTSISDSEHVDLLELYNLIDSHLLSVLPQ LEFKLSNDNEERLQVVKLLAKMFAGKASELASOKPAMQCYLGAFFNDILVNDILNF KFASHCLNHNHDLADLLEYLKVSHDPEEAIRHDYIVSYIYAAKDDILLVNDILNF VREPTLDRWRVRKPMAGMLAOIYKKVALQSAAGDAKQAJAMIKDLKLLHYONSID DLIVERIEFAOYVMPHNLFTNEMKCIYLYATATDLNPKVLANEMKCONLLRHQVD LLDLIKOPKTDASVKAITSKVAVYITRNLPDEKADDEPKKTOYLEDKTRKOLEVL VSPYSCSKQAEQCVETIRKLGPNKQPTNPPELMETKLELRIAPVHIDTESISALIKO VNKSIDGTADDEDEGVPTDQAIKAGLKLKYSFTNHPISFSAEFESLACLKMDDE KVAEALQIFKNTGSKIEDPPHRSALLPVLLHHSKRGPRQAKYAIHCIAHAFSSK EFOAFOIEPLHKSLEDPNLEHLITPLVTIGHIALADOPFAAPLKSIVAFETVLDL MNDRLRGKTKTKLWPDDEVSPEPMVKIOAIKMWVRMLGKNNHSGKSTRLITLT ILHSDGDLTEQKISKPDMSRLRLAGSAIYKLAQEPYHEITILEYQIOLALINDE CYOVRQVFAQKLHKLGLRLPLEYMAICALCANDPVERRAHAROCIVKNINAREY LKQNAVSEKLSLLEPYVPTVTHLHAHDYVYVODIEOLKDYKDECLMFLVETLMA KNENSHAFIRKMWENIKOTKDAQDPDDAKMKELTYCDVAMANIINSKSTYSLESP KDPVLPAREFTOPDKNFNTKNYLPEPEKSFETTPCKPKTYVYLGAVNKLPSAGKSO TKSSRMETVSNASSSSNSPGRJIKGRLDSEMSHSEMDYTMSSPLPGKSKDRDS DLVRSLEKPRGRKTPYTBQEKLGMDLTKYLVQEQPKSQSRKRGHARASEBEO QMPERKRLKEEDILENEDONSPPKGRGRPKPLGGGTPEEPTMKSTGSKSKGSP PPAPEEEEEEEOGNTQKSKSHRYSRRAOQAEPSSESALESTOSTQOKGRG SKTSPSQPKKRVYGRSKQATENDSSEEDVDFQSSPVDDIIPQEEFESEESTVN VRRRAKRRRR"
exon	/number=3 439..525 /number=4 526..623 /number=5 624..750 /number=6 751..831 /number=7 832..972 /number=8 973..1088 /number=9 1089..1183 /number=10 1184..1329 /number=11 1330..1481 /number=12 1482..1595 /number=13 1596..1677 /number=14 1678..1726 /number=15 1727..1866 /number=16 1866..2046 /note="matches EST AA687846 from clone IMAGE:1219676"
misc_feature	1866..1941 /note="matches EST AA687776 from clone IMAGE:1219628"
exon	1867..1982 /number=17 1983..2088 /number=18 2089..2249 /number=19 2250..2373
exon	















```

Qy 560 RQLEVLVS--PTCS-----CKAAGCVREITKKLGNPKQPTNPFLMIKIFLERIA 609
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 594 KSKNVAVSVEPTSSGVRSSRTLMKKDCGRLNKQVEKTRGN-----LRSLK 644
Qy 610 PVHDTESISALIKOVNKSIDGTADDEDE 638
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 645 ELNAETDRTA---EEQEVSLAESDDRSE 670

RESULT 14
Q9S9P0 PRELIMINARY; PRT; 990 AA.
ID Q9S9P0
AC Q9S9P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE T24D18.4.
GN T24D18.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A..
RC STRAIN=CV. COLUMBIA;
RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,
RA Toriumi M., Chin C., Chiu J., Choi E., Chung M., Gonzalez A.,
RA Hwang B., Koo T., Li J., Liu A., Vaysberg M., Altafi H., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,
RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
RA Davis R.W., Ecker J.R., Federlpiel N.A., Theologis A.;
RA "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A..
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RN Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A..
RC STRAIN=CV. COLUMBIA;
RA Theologis;
RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010924; AAF18491.1;
SQ SEQUENCE 990 AA; 109594 MW; 47854068D1B444FB CRC64;

Query Match 3.78; Score 264; DB 10; Length 990;
Best Local Similarity 17.5%; Pred. No. 7.6e-08;
Matches 241; Conservative 188; Mismatches 440; Indels 506; Gaps 50;

Qy 43 MDMDQSEERKELNLAHLASDFLLKHPCKDVRLLVACCLADIFRIYAPEAPYSPDK 102
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 47 VEQDLSSVQKALPPHRYALVSAD-LLNPPSDVRVSVVSCLTETMRITAPAPY-NDEQ 104
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 103 LKIDFMFTITRLKLGLEDTKSPQFNRYFYLLLENIAWKSYNICFELEDSNEIFTQLYRTLF 162
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 105 MKDIFQVITAEFAKELADASSRYKAEVILETVAKVRSSLVMDLE-CDDLVLVEMFORFL 163
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 163 SVINNGHOKVHMVMDLMSIIICEGTVSOELLDYVLVNLVPAHKNLKNQAYDLAKALL 222
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 164 KIIRPDHPQLVLVSMETIMITVIDESEVPMDLLEILLTTTVKQSDQVSPAALTILVEKVL 223
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 223 KRTAQAIPEYITTFNFVNLMLGKTSISLSEHVEDLILELYNIDSHLLSVLPOLF--- 279
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 224 SSCCKLQPCIM-----EALKSSGTS-----LDMS-----PVSSICSEFATT 263
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 280 -----KLKSDNDEERL---QV---KLAKMFGAKDSELASQNKPLMOCYLGRFNDIHVP 328
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 264 QAHNDVVKPDNEADEKISEGVVPVDSLEDKLNLGLSRKGRSRKSRGGRANGD--- 320
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

```

```

Qy 329 IRLECVKFASHCLMNHDPDLAKDLTEYLKVRSHDPPEATRHDVIVSIVTAARKDILLVNDH 388
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 321 ---EKVITANEGLSESTD-AETASGSTRKRGWKPKSLMNPBEGYSFKTSSSKV----- 370
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 389 LLNFVRENTLDRWRVRKEAMMGLAQIYKVALOSAACKDAKQIAWIKDLKLLHIYQNS 448
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 371 -----QEKELG-----DSSLGKVAARKVP----- 389
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 449 IDRLLLVERIFAQYVMVPHNLETTERMKCLYYLYATLDLNAVKALNMWKKCNQLRHQVKD 508
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 390 -----LPSKVGQTNQ-----SVVISLSSSGRARTGSRKRSR- 420
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 509 LLDLIKQPKTDASVKAIFSKVMVITRNLPDQKGAQDFMKKFTQVLEDDK---IRKQLEVL 566
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 421 ----TKMEETHDVSSVATQ-----PAKKQT-VKKNTPAKEDLTKSNVKKHEDGI 465
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 567 VSPTCSCQA--EGCVREITKKLGNPKQPTNPFLF--MIKELLERIAPVHIDTESISALI 622
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 466 KTGSSKKEKADNGLAKTSARK-----PLAETMMVKPSGKKL--VHSDAK----- 508
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 623 KOVNSIDGTADDEDEGVPTDQAIAGLELLKVLSTFTHPISFHSAETFESLIACLKMDDE 682
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 509 --KKNSEGASMD---TPIQSSKS-----KKKDS 532
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 583 KVAEAAALQIFKNTGSKTEEDPHIRTSALLPVLHHSKKGPPROAKYAIHCHIAIFSSKET 742
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 533 RATTPTAK-----KSEQAKSHPKMKR-----IAGEEV 560
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 743 QFAQIFELHLKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPKWSWATFIVKDLLMNDR 802
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 561 E-----SNTNEL-----GEE 570
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 803 LPGKTTKLWYDDEVSPEPMVKIQAIKMVRWLLGMKNHNSKSGTSLRLTLTLHSDG 862
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 571 LVGRVNVNWPDLKKFEGVKSVCVRVKM-----HQ-----VTYSDG 608
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 863 DLTEQKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLEQYQLCALAINDECYQVRQVF 922
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 609 DV-BELNLKK---ERFKIIEKSSASEDKEDDLLESTPLSAF----- 646
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 923 AQLHKLGLSLRLPLEYMAICALCAKADPDKERRAHARQCLVKNINVRREYLKQHAASEK 982
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 647 -----SSPEVRSSMOTMKKD-----IQREKSKKRKIVSKN 661
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 983 LLSLLPEYVVPYTIHLAHPDYVKVQDIEQLKDYKCELFVLEILMAKNENNENSHAFIRK 1042
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 662 -----VEP-----SSPEVRSSMOTMKKD-----S 682
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1043 MVENIKQTKDAQGPDADAKMNEKLYTVCVAMNINMSKSTTYSLESKPDVLPAREFTQPD 1102
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 683 VTDSIKQTKRTKGALKAVSNP-----ESTTGKNLKSLE-----KLNGEPD 723
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1103 KNFSNTKNYLPENKMSFTPGPKTNNVLGAVNKPPLSAGKOSOTKSSRMETVSNASSS 1162
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 724 KTRGT-----CKQ-----KVTQAMHRKIEK----- 745
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1163 NPSPGRIKRLDSSMDHSENYDTMSSPLPGKSKDRDSDLVSRSELEKPRGRKTPV 1222
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 746 -----DCDEQEDLETKEEDSLK-LKESDA--EPDRMEDHQELPENHNVTETK 790
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1223 TEOREKLGMDLTLKLVQOKPKGQSRSPKRGHTASEDEQOWPEKRLKEDILE--NEDE 1280
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 791 TDGEE-----QEAKEPTAESKTNGEENAEPTDCKEHEKSKLKEPNAEPKSDGE 839
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1281 QNSPPKK-----GKRGRPPKPLGGTTPKEEPTMTKS-----KKGSKKSGPPAPEEE 1328
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 840 EOEAAKEPNAELKTDGNOEAKEKELTAERKTDDEEHKVADEVEOKSQKXETNVEPAESEE 899
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Qy 1329 EEEQSGNTEQKSKSHOVRSRRAQQRAESPESSEIAESTQSTPQKGRGRPSKTPS 1383
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
Db 900 QKSVEEPNAEPKTK-----VEEKESAKEDQATDKLIEKEDMSKTKGEBIDKETVS 949
:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

```



```
O94237
ID O94237 PRELIMINARY; PRT; 390 AA.
AC O94237;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)
DE HYPOTHETICAL 45.2 KDA PROTEIN (FRAGMENT).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee M., Yoo H.S., Chung K.S.;
RT "clone 17 (bimD homologue).";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049529; AAD02493.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 390 AA; 45161 MW; ECC83FE5A45CA8C6 CRC64;

Query Match 4.3%; Score 309.5; DB 3; Length 390;
Best Local Similarity 24.8%; Pred. No. 3.5e-11;
Matches 105; Conservative 76; Mismatches 192; Indels 51; Gaps 11;

QY 825 KTOATKMYRVLGKNNHNS--KSTSTLRLTLTILSHDGLTEGKISKPDMSRLRLAA 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 KVLAIKLVNRLRAAAGGTALNIGAPIIKLLKLVLLMADGELSPFKNTPKISRVLRLTA 63

QY 883 GSAIVKLAQPCYHEITLLEQYQLCALINDECYQVQVFAQKHLKGLSLRLPLEYMAI 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 SKYFLKLSIPYABHIDFSSVSNLLIMPENFDVRLNLTQLKQKQLKPLISYVPL 123

QY 943 CALCARDPVKERRAHARQCLVNINVRRLKOHAAVSEKLLSLPEYVVPYTIHLLAH 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 LFLTAVDPEELKTKA-----SWIRSQVAFQKTHDFTMEYVATYLIHLLSH 172

QY 1003 PDKYVQDTEQLKVKECMLFVLEILMAKNENNSHAFIRKMYENIKQTKDAQGPDADKN 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 PDISTESSENSLDFATIRYFVTVVNSNV---PIVFLMQRIKQSYDV---IEDG--N 224

QY 1063 EKLTVCDVAMNIMSKS-----TTY--SLESPKDPVLPARFTQDPKNSFNKNYLP 1113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 NIIYVLSDMAQKILQVKSQNFQNSLTYPKQIKLPEILRIPSIDKKRIEN--KIFT 282

QY 1114 PEMKSFPTPGKPTNVLGAVNKLSSAGKQSQTKSRMETVSNSSNPNPSSPGRIKGR 1173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 PRMES-----QIEHAIRTPVSSFAKQTTNKHANLKQKTHSSKSDKSSRRRK-- 330

QY 1174 LOSSEMDHSENYTMSPLGKSKDRDDSLVRSELEKPRGRKTPVTEQEEKLGMD 1233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 ---NEKRRKLNEQNPNIRNVPRSSSRFQIRINYSAPS---SSEETSEETSEED 383

QY 1234 LTKL 1237
Db 384 FDEI 387

RESULT 13
Q9SA10
ID Q9SA10 PRELIMINARY; PRT; 780 AA.
AC Q9SA10;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE F23A5.16 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S., Lee J.M., Li J.,
RA Gonzalez A., Liu A., Sakano H., Koo T., Pham P., Vaysberg M.,
RA Hwang B., Chin C., Choi E., Chiou J., Altafi H., Brooks S., Chao Q.,
RA Conn L., Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C.,
RA Lam B., Nguyen M., Palm C., Shinn P., Tambunga G., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F23A5 sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Theologis;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011713; AAF14668.1; -.
SQ SEQUENCE 780 AA; 88256 MW; 8BDE2E90DCDDF22B CRC64;

Query Match 4.2%; Score 303; DB 10; Length 780;
Best Local Similarity 20.4%; Pred. No. 2.3e-10;
Matches 153; Conservative 118; Mismatches 254; Indels 224; Gaps 27;

QY 18 GVKETSDKISKEEMVRLKMWVKTMDMDDEEKEKLYLNAL----HLASDFLKH 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 GNLNLSPESSDOLLTLDETESLLKNVEQD---OPLSMQSAALIPSRNALVSVLLSH 73

QY 73 GKDVRLVACCLADIFRIYAPAPYTPSPDKLKDIFMTITRQLKGLDTSKSPFNRYFYLL 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 DSDVRVSVVSCLTIVRITAPETPY--SDDLMEIFRLTIEAFEKLADASSRYKKAEPVL 132

QY 133 ENIAWKSYNICFELEDSNEIFTQYRLFVINGNHOKVHMVMDLMSIIICEDGTVS 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 DNVAKVSKCLVMDLE-CYDLILOMFRNFFKIRSDHPQLQVFSMELIMIALIDEQVS 191

QY 193 QELLTVLVNLPANLKNQAYDLAKLLKRTAQAIEPYITTFNOVLMCLKTSDLS 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 TDLLOSILATVKEKQNVSPMSWLAELVLSRCARKLKPYII----EALKSRGTSLDWVS 247

QY 253 EHVFLILELYNIDSHLLSVLPQLFEFLKLSNDNEERLQV-----VKLLAK----- 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 PVVSSICQSVFN-----TPKVHSPVNTKEHESLKQVRSSTDAEITGRGRKPNSL 298

QY 299 -----MEGAKDSELASQNKPLWQ-----CYLGRFNDIHPIRLECVKFASHCLM 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 MNPEDYDISWLSGKRDPLKTSNKKIQKGGVSSLGKVPAPKKTPLPKE----- 348

QY 343 NHP-DLAKDLTEYVK-----VRSHDPEE----- 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 NSPATSSRLTSLKRSRVKMDSDYSDLSLSPRLKKLASCDFDEEPNQEDDRKIGNSS 408

QY 365 ----AIRHDVIVSIVTAAKKDIIL---LVND-----HLLNFVRERTLDKRWVRKEAMGLA 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 KQTRSKNGLEKSKQTKAKKPPVVEAKIVNSSGKRLSARSVAKRRNLE---RAPLDTLVPS 465

QY 414 QIYKVALQSA-----GKDAKQ-----IAWIKR 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 SRKKMVSQVAARQLANESSEETPKSHPTRRRTVRKEVESDGFEDLVGKRVNIMWPLDK 525

QY 440 LLHIYQNSIDRLLVERIFAQYVMVPHNLETTERMKCLVYLATLDLNAVKAALNEMWCKQ 499
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 ---TFEGVID-----SYCTRKKMHRVIVSDGSEELNTEERW--- 561

QY 500 NLLRHQVQKLLDLIKPKPTDASVKAIFSKVMVITRNLPDGPQAQDPKPKTQVLEDEKI 559
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 -----ELLEDDTSADEKKEI-----DLPEISPLSDIMQR-----QKV 593
```



```
Db 61 GRPKPLGGTPKPEPTMKTSKSKGSGPPAPPEEEEEERQSGNTEQSKSKQHRVSR 120
Qy 1351 RAQR 1355
Db 121 RAQR 125

RESULT 11
ID 094076 PRELIMINARY; PRT; 1506 AA.
AC 094076;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE BMD PROTEIN.
GN BMD.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93387663; PubMed=8375649;
RA Denison S.H., Kaefer E., May G.S.;
RT "Mutation in the bmd gene of Aspergillus nidulans confers a
RL conditional mitotic block and sensitivity to DNA damaging agents.";
RL Genetics 134:1085-1096(1993).
CC -!- FUNCTION: REQUIRED FOR PROGRESSION THROUGH MITOSIS AND ALSO PLAYS
CC A ROLE IN REPAIR OF DNA DAMAGE.
CC -!- SUBCELLULAR LOCATION: KINETOCORE OR MITOTIC SPINDLE (POTENTIAL).
CC -!- SIMILARITY: SOME, TO B-ZIP PROTEINS.
DR EMBL; L03200; AA03063.1; -
KW Cell cycle; Mitosis; DNA-binding; DNA repair.
SQ SEQUENCE 1506 AA; 166270 MW; F50B0893EE767579 CRC64;

Query Match 8.88; Score 632.5; DB 3; Length 1506;
Best Local Similarity 19.98; Pred. No. 4.5e-30;
Matches 308; Conservative 298; Mismatches 626; Indels 319; Gaps 53;

Qy 30 EMVRLKVVVTFMDMDQDSEEEKLYLNLAHLASDFLKHGKDVRLLVACCLADIFR 89
Db 47 DLLQRLQTLAQLPNVEQF-EIEKESLRKVSOELATAQLLAHKDGVRAWATCCIVDLR 105
Qy 90 IYAPEAPYTPDKLKDIFMTIRO-LKGLDTKSPQFNRYFYLLNIAMVSKSYNTCFELE 148
Db 106 LCAPDAPETA-NQLKDIFTCIVSSIIIPALGDPSPNNAQHIYVLSNLAEVKASIVLMTDLD 164
Qy 149 DSNEIFTOLYTLFSVINNGHN-----QKVHMHVYDLMASSIIICEGDTVSQELLDTVL 200
Db 165 HPOTLIVPLFISCFDIVAGSAKASTGEPVAKNVEYDMTRLVTVIDESPVLADPVVIV 224
Qy 201 VNLV-----PAKKNLNKQ-----AYDLAKALLKRTAQAEIYIT 234
Db 225 AQFLRVDPRLVDGCKKCKKPEQVDEKQETLLKDYPAAYNMAKICQACPERTSHS 284
Qy 235 TFFNOVL-----LGTSTISDSEHFV-----LILELYNIDSHLLL 271
Db 285 QYFNNVIIDASATGTQGPSQARRTNLDDSEGEDIKELSKAHLRIELWRACPDVLQ 344
Qy 272 SVLPQLEPKASNDNEERLOVQLLAKMFGAKDSLASONKPL----- 314
Db 345 NVIPOLEAELSAESVSLRLATOTIGDL--TSGTGAGVAGPPPLPMDPVPVQVQLDDYAR 402
Qy 315 -----WQCYLGRFNDITHVIRLECVKFAASHCLMN----- 343
Db 403 SIPQPNVLLMPFAPKPFQASHSAVDSFLSRLDKSASVRSWRTAIGRIILTSGGSL 462
Qy 344 HPDLAKDITELYKVSHPDEBAIRHDVITSVTAKKDI-----LLVNDHLLNVR 394
Db 463 SDNEQOTLITHLSMLRDADERVRLAAVEAVGTGFLGSHVKNKLGVSQVSTQDSLLFLA 522
Qy 395 ERTLDKRWVRKEAMWGLAQIYKYALQSAAGKAAQIAWI-----KDKLLHIYQNSID 450
```

```
Db 523 ERVDRKSQVREHATKYL-----RAMAVASGDIERSHQVTPLLKEAPSRILDAYTNDPE 579
Qy 451 DRLIVERIFAQYMP-----HNETTE-----RMKCLY 479
Db 580 IHVSIDRAMPEILLPSYPIKPLSRSSSSQSLKDSQAAPESADVDRIVRILT 639
Qy 480 LYATLDLNAVKALNEMKQCNLLRHQVLDLLIKQ-----PKTDASVKAIFSVM-VI 532
Db 640 LVGGLDEKAKKVFAMQKROVSLRTAVTVYLQACEEYNGVMENKDKQIKAKLTIVDAL 699
Qy 533 TRNLPDPCGAQDFMKKFTQVLEDEKIRKQLEVLVSPTCCKQABEGCVREITKTLGNPKQ 592
Db 700 AKTFPDPAITSADLWKFAKI--HRRGYQLIRFAMAAVSDYRTVIKIKELARRL---OS 754
Qy 593 PTNPFL-EMIKFLERIAVPHIDTESALIKQVKNKSDGTADDEDEGVPTDQAIRAGLE 651
Db 755 SNNILHETLTLLYRCSSIVFNRSHPA-IMSISRS-----DENGLAAPAH-----E 801
Qy 652 LLKVLSTFHPISFHSATFESLL--ACUKMD-----DEKVAEAAALQIFRWNG 696
Db 802 MLKEISSLP-----EVLEAQVQEIICKLEAQAPKATTVSAAAGTEELKACSGPAKKLP 855
Qy 697 SKIEDPFIHSALLPVLHHSKKGPPROAKYAIHICHAIFSSKETQFAQIFELHKSLD 756
Db 856 SKLPKERKFFQALVDYALHSPS-----PRAAKHAVLILMAV-TDKKNMYAKDLVEKVCSC 910
Qy 757 PSNLEHLTPIVTVIGHIALLAPDOFAAPKSWATFIVKOLLMDRLPGKTKTKLWVDE 816
Db 911 TYDSERFLTATLSQNLAPRE-ADESDAIKISVNOILLTNRSPTPNSGYFW--SD 967
Qy 817 EVSPETWVKIOAIKMMVRLMGMKNHKSSTST-----LRLTTILHSDGDLT 865
Db 968 QVDDETAKEWALKIIVNRL-----RAKDGSDSDDFRAHAEPVYDTLNLKLVNSGELS 1021
Qy 866 EGGKISKPDMSRLLAGSAIVKL-AQEPYHEIITLQYQOLCALAINDECYQVROVFAQ 924
Db 1022 KKQTPATQKRLRLAANSLLKLCSSHALCEQLLTPODFNSIALVAQDPLPEVRSGETN 1081
Qy 925 KLHGLSR-LRLPLEYMAICALCARDPVKERRAHARQCLVKNINVRREYLKQHA----- 977
Db 1082 QLKKKLVQDTRLGARWVVIYLLAFEP---QVGLKQDSTLWLSRAAFPSQOTNGKKE 1137
Qy 978 --AVSEKLLSLPEVVPYTHLLAHPDY--VKVODTEQLKDVKECLWFVLEILMARNE 1033
Db 1138 KQTYMEALFSRL-----LSLLAYHPDYPADLDESTKLDLTDFAFYILFYSAVAN 1189
Qy 1034 NNSHAFIRKMWENIKQTKDAOGPDQAKMNEKLYTVCDVAMNITNSKTYTSLSPKDPVL 1093
Db 1190 EHNLSLPHIAQVRKQARDGTIKSD-EMSRRLHTLSDLAQATIRRFADVYQQ----- 1241
Qy 1094 PARFTQDPKNSNTKNYLPPEMKSFPTPGKPTTNVLGAVNKPLSSAGKQSQTKSS--- 1150
Db 1242 --RRFGGAGGVNLLQTY-----PGK--VGVPSSIFAPMGSHREAOEADVTFL 1286
Qy 1151 -----RMETVSNAS-SSSNPSPGRINGRDSSMDHSENEEDYTMSSPLCKKSKDR-D 1202
Db 1287 PEDAEKRLDRIVRATMTKNGSQAGAKRKRTDSTQEPSRDANAARKKRNSESSSRKS 1346
Qy 1203 DSDLVRSLEKPRGKKT--PVTEQEEKLGMDLTKLVQEQPKGQSRKRGHTASESD 1260
Db 1347 GSSAVGFKMPKRSKTKYGDWSSDGEAGNVAASSATRRRSNRGCS-ASRISYADPDS 1405
Qy 1261 E--QWPEEKRLKEDIENEDEQNSPPKKGRGRPPKPLGGTTPKEEPTMTKTSKGSKKK 1318
Db 1406 EDDMEMDELNAQRODEDEGQ-----AKDIENGSDLSE---LSEADSNWL 1448
Qy 1319 SGP-----PAPEEEEERQSGNTEQSKS-----KQHRVSRRAQRAESP 1359
Db 1449 BEPEDDDGPSEKEDQDDKQDDDDAQAPSPVPVASKAKVPCKAMKATLP 1499
```

RESULT 12



```

Db 33 LTRPG-----KPIVDTLKRLDALSKELSDMDQETVDTDSL-VKAVKDVASHQIOLKH 86
QY 73 GKDVRLLVACCLADIPRIYAPAPYSPDKLKDIF-MFITROLKGLDTSKPOFNRYFYL 131
Db 87 DKGVRATACCIIDILRLCAPAPFT-PSOLQDIFNLINSIIPALFNSPNYNNHOKVY 145
QY 132 LENIAVKSYNICFELEDNEIFTOLYRLFVSINNGHOK-----VHMIMVDMLSI 184
Db 146 LRSFAIKSIVILLDVEGSEALLKLTFTIFDGVSLKSKSGEOGVKDVFFSMQEMIGAL 205
QY 185 ICGDVTVOELDTVLVN-LVPAHKNLNQAYD-----LAKALLKRTAQAI 230
Db 206 IDSVTLPKGVVDVINAQFLRAAAPGLGERQDHVPIDDSQATLLKKEPEAYQMVRNLC 265
QY 231 -----PYITFFNOVLM-----LGK-----TSISLSE-----H 254
Db 266 QYDDKMAFASQYSDVIVDATGFAKPGNSRDEDEGDPGTPGSESDLKELRAH 325
QY 255 VFOLILELYNIDSHLLSVLPOLFEFLKSNDEERLQVVKLLAKM-----FGAK----- 303
Db 326 V--LIREIKWAPMILQVWPQVDAELSDNVLHROMATETIGDMISGIGACAGPPPLPIL 383
QY 304 -----DSELASQKPLWCYLGFRFNDIHPVIRLECV 334
Db 384 DPAAVPLSLEEDRAEPVVTNLTPLCLCSISFSQTHSTTFHFLSKNDKAPSIRAAWT 443
QY 335 KFASHCLM-----NHPDLAKDLEYLKVSRHDPPEAIRHDVIVSVIYTAACKDIL 383
Db 444 TAVGHILSTSAQIGLSRDEATLIRGLGELS-----DSDEKVLRSVAKAVETFRFQDII 499
QY 384 L-----VNDHLLNFVRETRDKRVRKEMGLAQIYKVALQSAAGDA-AKOI 433
Db 500 AKLPGNGVGKDGSVLNTLADRCRDKRPVAVVAAMSELLAKLWAGTGMAGNEAVTAAL 559
QY 434 AMIKDKLLHIYQNSIDRDLVERIPAQYWP----- 465
Db 560 SGIPSIYNAFYANDLELVMDRVIEFLVPLGPPPAKATRNNSNANGNSQSASANAAS 619
QY 466 --HNLETTERMKLYLYATLDLNAVKALENNMKCONILLRHQVKDLDDL-----IKOP 516
Db 620 IDHDAITRAER---ILLARSLEPAKAFKAFQMSRRPQFAKILETYLDQCDRYNGVMS 676
QY 517 KTDASVKAIFKVMVTRNLPDQKADQFMKFTQVLEDDKIRKOLEVLVSPTCSCKQA 576
Db 677 NADKITSLNKADYTAQPLPHEVSKTDLIFAKI--HRRNRYNLKIVIGENDFKIV 734
QY 577 EGCVRITKGLGNPKOPTNPFLEMIKFLERAPVHIDFESISALIKOVNKSIDGTADDE 636
Db 735 YKALKELIKRMAKSDPS--VIDTLLPLLYRSGCLLFRNLSHLSITIMEYSKSKDKGLGS-- 790
QY 637 DEGVPDQAIRAGLELLKVLSTFHPISF--HSAETTESLL-----ACLKMDDEKVAEALQ 690
Db 791 -----TAHETLNEISQNPDLFTHTIGQLCKDLVDQAPTATKPNDPVIAV-TLK 838
QY 691 IFKNTGSKIEEDPPHRSALLPVLHHSKKGPPROAKYAIHCIAIFSSKETQFA-OIFE 749
Db 839 ACSTYARKPKDVMDRKFQVOTWINALYQVPVASKHAVNIVLCQDKDSWATDILQ 898
QY 750 PLHK--SLDPSNLHLITPLVTIGHIALLAPDQFAAPKWSWATFIVKOLLMDRLPGKR 807
Db 899 RILKGSYGSNN---FLNKLTAVSQLELLAP-KVTEDASDEILNLTFLKQLQVRTDAKD 954
QY 808 TTKLWVPDEVSPETMVKTOAKMM---VRLLGMKNHNSKSTSTLRLTLILHSDGL 864
Db 955 SDPDWYNGADMDDEIQAKLSLILVNRVRSIEGIEAKEA-SSVMKVLRIIEKEGI 1013
QY 865 TEOGKISKPMRSLRAGSAIVKLAQEPCEYHEIITLEOYLQCALAINDECQVROVFAQ 924
Db 1014 VEKETPKHKHTRLLRLAQLMLKICTQKHFDMDLTPSFNLLALTQDMVEVRHGTVR 1073
QY 925 KLHGSLRLPLLEYMAICALCAKDPVKERRAHAROCVLKNTNVRREYLRKQHAASEKLL 984

```

```

Db 1074 KLQKYLADGKLRSRYTYIIFLTAPEPSVEFK-----NRVETWISRARHFQNLK 1122
QY 985 SLLPEVVVYPTTHLAHDPDYKVKVDIEOLKDVKECLWFVLEITLMAKNENNSHAFIRKMV 1044
Db 1123 QVLEIMARLISLLAHPDY--SNEVEYLVHAR---YILNTVILVGTESNLGLYKYA 1177
QY 1045 ENIKOTKDAQGPDDAKANEKLYTVCDAVNII---MSKSTTYSLESKPDVLPARFET-- 1099
Db 1178 ERVKOTQDGLNPN---SDAHRVLSDLAQSIIRKWOEKVWFAAFPGKVGLPGLLYTAL 1233
QY 1100 --OPDKNFNTKNYLPP-----EMKSFPTPGPKTNTVNLGAVNKLPSAG 1142
Db 1234 QSHSEAQAIAEKSYLPEGLDEKLDELLRAMDRKKRKSTSAATGTNNRDCHKSKVKSSG 1293
QY 1143 KQS---QTKSSRMETVSNASSSNPSPGRIKGRLSSEMDHSENEEDYTMSSPLPKCKSD 1199
Db 1294 DRDGDREAKSSR--KIATAKSKKTPAK-----KRSRDODWSPGAAPV--ESE 1336
QY 1200 KRDDSDLVRSLEKPRGRKKTVPTEQE-----EKLGMDDLTKLVQVQPK 1244
Db 1337 RQSS-----RRAANKAKGPTYTERDSDDDDDEMLEGVAWEYLEDDEDKEDDEGS 1389
QY 1245 GSO-----RSKRKHTASED-EQOWPEERKRLKEDILENEDE-----QN 1282
Db 1390 GEEAPRPTAAKASATTRSSRREDKESEBEVNEEPEKEQEEDADPEQEEAPTAAAKRT 1449
QY 1283 SPKPKGR-----GRPKPLGGCTPKEEPTMKTSKGSKKSGPPAP 1324
Db 1450 SATAKGKAASLPVRNAAASTRRSGRGKAAASAKDKDDSDSELSDVMVDEPEA 1509
QY 1325 E-----EEEEEROSGNTQESKSKQHRVSRRAQRAESPESSAIESTQST 1370
Db 1510 EAEAEAEAEAEVEEVEDEKEQEQAAPPARANGRKAAPAKAAAPKGAKAADASA 1569
QY 1371 POKGRGRPSK 1380
Db 1570 PARRSGRSTR 1579

RESULT 10
Q9H5N8 PRELIMINARY; PRT; 125 AA.
AC Q9H5N8;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CDNA: FLJ23236 FIS, CLONE COL00725.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RA "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026889; BAB15584.1;
SQ SEQUENCE 125 AA; 14281 MW; 492809836D3665D3 CRC64;

Query Match 9.1%; Score 656; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. NO. 5.4e-33;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 MDLTKLVQEQPKGSQSRKRGHTASEDEQOWPEERKRLKEDILENEDEONSPPKGKR 1290
Db 1 MDLTKLVQEQPKGSQSRKRGHTASEDEQOWPEERKRLKEDILENEDEONSPPKGKR 60
QY 1291 GRPPKPLGGCTPKEEPTMKTSKGSKKSGPPAPEEEEEEROSGNTQESKSKQHRVSR 1350

```



Query Match 11.2%; Score 807; DB 5; Length 1579;  
Best Local Similarity 21.5%; Pred. No. 1.1e-40;  
Matches 352; Conservative 282; Mismatches 674; Indels 326; Gaps 57;

Qy 11 GKITYPGVKESDKISKEEMVRRLKVVKTFTMDMDQDEEE---KELYLNALHLASDF 67  
Db 3 GDVYPOGCLPILNSNNVTOQIERLURKLFNCHDKCTNNESEVSPNRFARLQHLSEOC 62

Qy 68 FLKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLIDFMFITROLKLEDT--KSPQF 125  
Db 63 FLDNSNTDFRILLSLCLANILRIFQPELPTSPVMDKEVYVLFRTMRGLGVDVTDSPKF 122

Qy 126 NRYFYLLENIAWVKYNICFELEDSENE-----IFTOLYRTLFVSV-INNGHQ----- 171  
Db 123 KNYSLSVETME--KIIPPIIEMKHDDKEATPVFRALIKLILAPCGKGNQNKKEARL 180

Qy 172 -----KVHMHVMDLMSSTICGDTVSQBELLTVLVNLVPAHKN 209  
Db 181 LKIQDNDSDNDDEEDENAEKIRKSLIQATVITNLDVQNECLDVLFYHIINPORS 240

Qy 210 LNKQAYDLAKALK-----RTAQAIPEYITTFNFQVLMGK-----TSISDLSEH 254  
Db 241 NFAEARALAEIDIIRCSDNESDTLANSIRSTWTAAAKE-----GKLPEEFELTGSSNRSK- 295

Qy 255 VFDLILELYNDSHLLSVLPQLEPKLSNDNEERLQVVKLLAKMFGAKDSELA--SQNKP 313  
Db 296 FFEVLYRHYVSFVLVSGAIOELKFWLQSENEQYRKEAVTVVGMLTRDKHCFQFGMDSDNP 355

Qy 314 LWQCVLGRFNDTHVPRLCEVKFASHCLM--NHPDLAKDLTEVLRVSHDPEAIRHDVIV 372  
Db 356 TWSAFLNASIDQDSDVRHEFFVQSKDILISNHSRLRGOIINSRLSLVDLDDIRRDVVT 415

Qy 373 SIVTAACKDILLVNDHLLNFVRERTLDRWRVRKBEAMGLAQIYKYLQ----- 422  
Db 416 GVTEVAKTKLEVISDKMLCAEAKMRKPKVRIQAIKRLMDLYNHVMTSSPPQPFSDKG 475

Qy 423 SAAGKDA-----AKQIAWIKDLHLHY-----YNSIDRLLVERIFAQYM 463  
Db 476 SCAPKSKSEATLSYTESKESVRFISTAVFNRYLRTORLPAYH-----DARATIERFYQLY 532

Qy 464 VPHNLLETTERMKCLYLYATDLNVAKALNEMWKQNLRLHQVKDLDLIKQ--PKTDAS 521  
Db 533 VPYKVEPLRLVRLMADLFRNLDDICMLFGDIINSSQLRRAMIGILSQVQFTWMTQS 592

Qy 522 VKAIFSKVWVITRNLPDGGKAQDFMKPE--TVLEDEK---IRKOLEVLVSPSCCKQAE 577  
Db 593 SAQLKERIRRICQIPDAQILEKNMVFVQMAENDETENLVKMLSESYTSEENARTAG 652

Qy 578 GCVREITKLGKPNQPTNPFLEMIFKELLERAPVHIDTESISALIKQVNSIDGTADDED 637  
Db 653 ALQATMEKSKSKAQOT-----VFRHFIDRIVPLSFVDVPTAKEIHLVSDTVCAKVDLKK 707

Qy 638 EGVPTDQAIRAGLELLKVLSTHPTISFSAETFESLLACLKMDDEKVA--EAAL-----QIF 692  
Db 708 W---AENCFEXDLSLKIFTDNFGVLFDAEQIEIERSKILASEEPIAEALHVLSKIF 764

Qy 693 KNTGSKIEEDPHIRSLALLPVLHHSK-----KGPP---RQAKYATHCTHA 735  
Db 765 ANS-----HFRNKLENEATHKEKFWLLGLGKDLKDLVMREPELRSSCKLATRLLSF 815

Qy 736 IFSSKETQFAIFELPHLSLQPSNLEHLITPLVTIGHTALLAPDOFAAPKWSWATFI-- 793  
Db 816 ILGKEKV--IEFFD-----DQIEQLISRLYIESOGAANAQVGLGEIFRCDISYILPQ 865

Qy 794 VKDLMLNDRL--PGKKTTLKLVDPDEVS-----PE--TWVKIOAKIMVFWLLGCM 839  
Db 866 VMDVVESEKIGPMILTSSMHGNDOPVEFNEMLHTEKQWPKYAMAKVYAANFATKVLIVY 925

Qy 840 -----KNHNSKSGTSTLRLLTLTILHSDGLTBDQGKSKPDMSRLRLAAGSATVILQAO 891  
Db 926 PLIPSTEDKRMEKAAQNFIDLLSEIEKKGDG--GGQCDCEQARLRATASGCLLKLAS 983

Qy 892 EPCYHEITLQYQALCAINDEBCYQVQVFAQKLHKLHGLSRLRLPLEYMAI-----CALC 946

Db 984 VITYRTKLINTHFKNMSYIITDEAYCVRLYYALHVKKGLSKNRLPIEFAACYGLVNLGLS 1043  
Qy 947 AKDPKERRAHAROCVKNINVRREYLKQHAASVEKLLSL-----LPEVVVPYTHL 998  
Db 1044 EEDGENKMDGFEKTCM---NQAQAFGERNDNQATLLKLEGQRAIFCSSEVIAVYVWL 1099  
Qy 999 LAH-----DPDYVKVQDIEBOLKDVKECLMFVLEILMAKNENNSHAFIR 1041  
Db 1100 LANYDKLEKVEGNANRNDSSSELEIKVANVNLSELQSLWLVIDSLKIAKCMQKVW-- 1157  
Qy 1042 KMWENIKOTKDAQPDPAKM-----NEKLYTVCDVAMNIM----- 1077  
Db 1158 KVLKELKTCGDKSMRSDSLSTTLREHNEVNPDSGRNIKWLVCGLGITMMLYRAKLOM 1217  
Qy 1078 ---SKSTYSLE-----SPKDPVLPARFFTOPDKNFNTKN-----YLPPEMK 1117  
Db 1218 EDQEAKEAGFNLPFYVCSPKDKADPSNVA--PDVLIENKRNRIKPKGRVPHVSDLT 1276  
Qy 1118 SFFTPGPKPTTNVLGAVNKPLSSAGKQKOTKSSRMETVSNASSSNPSSPGRIKRLDSS 1177  
Db 1277 SEFTP--PPQGN-----ESTGSNSKNASRRANVTGTRKRGAGTKTKRKSOGS 1325  
Qy 1178 EMDHSENEGYTWSSPLPGKKSDKRDSDLVRSLEK-----PRGRKKTPVTEQEKLCM 1231  
Db 1326 KISDDEGSDMDVKSPAVKTKRSKRGEDLPDEEDEMEVILPKRRGAAPDSTIASSSN 1385  
Qy 1232 DDLTKLVQEQKP-----KGSQSRKRGHTASEDEQOQWPEEKRLKEDILENEDQN 1282  
Db 1386 GSILKNGSGSPKKRNSRGVSKGRONSTKEAIFDVSDDEM-----EETKRDVNS 1437  
Qy 1283 -----SP---PKGKRGRRPKPLGGGT--PKEPTMTKSKGSKKSGPPAPEEEEE 1331  
Db 1438 LDNLIISILDSESSGTRRSARTIATATITSSTPLVTPKPKITRKRSEAVVEVEDEE 1497  
Qy 1332 ROSGNTQKSKSKQHRVSR-----AQORAESPESSAI--ESTOSTPKGR--GAP- 1378  
Db 1498 PEI--NDKRSPPKRVSGRRSAPTPTKNSKAPVSPKKTKLTKETNGVSPKKNYGLPM 1555  
Qy 1379 -----SKTPSPS 1385  
Db 1556 BEDDDSIGKTPKPT 1569

RESULT 9  
QYUVY6  
ID QYUVY6 PRELIMINARY; PRT; 1596 AA.  
AC QYUVY6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE SPO76 PROTEIN.  
GN SPO76.  
OS Sordaria macrospora.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Sordaria.  
OX NCBI\_TaxID=5147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Heemst D., Poggeler S., Zickler D.;  
RT "The SPO76/PDS5/BLM gene involved in mitotic sister chromatid  
RT cohesion is also needed for synaptonemal complex formation and  
RT recombination in Sordaria macrospora.";  
RL Submitter (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ009934; CAB51808.1; -;  
SQ SEQUENCE 1596 AA; 176745 MW; B03FBB8C3E63AEC2 CRC64;

Query Match 10.4%; Score 747.5; DB 3; Length 1596;  
Best Local Similarity 20.7%; Pred. No. 4.7e-37;  
Matches 338; Conservative 295; Mismatches 652; Indels 345; Gaps 51;  
Qy 13 ITPPGVKEISDKISKEEMVRRLKVVKTFTMDMDQDSEEEKELYLNALHLASDFLAKHP 72











Q9V646;  
 01-MAY-2000 (TrEMBLrel. 13, Created)  
 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 CG17509 PROTEIN.  
 GN CG17509.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,  
 RA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR ENBL; AE003823; AAF58593.1; -;  
 DR FlyBase; FBgn0033676; CG17509.  
 SQ SEQUENCE 1218 AA; 138887 MW; F4A6B7238A0801C2 CRC64;

Query Match 27.6%; Score 1982.5; DB 5; Length 1218;  
 Best Local Similarity 37.3%; Pred. No. 1.5e-112;  
 Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;  
 13 ITVPPGVKESDKISKEMVRRLKVVVKTMDMDQDSEEEKLYLNALHLASDFELKHP 72  
 4 IVTPTGCRPLVEDGTDLIRRLUKTLANVLQTDQDNLQYQ-YIPALHLDDDFMQHP 62  
 73 GKDVRLVACCLADIFRIAPAYTSPDKIDIFITRLQKGLDTPQPNRYFYLL 132  
 63 SRDQVLLIACCVADVLRVAPAYPEAYKEQDIKIFEFIKQLHGLKDRPDSKRYFYLL 122  
 133 ENTAWKSYNICELEDSNEIFTQLYRTLSVINNGHNQKVMHMDLMSIICEGTVS 192  
 123 ENLAFVKSFNMCFELEDQCFQDFLSTIFKIYVDQHSVKVTNFFLDVLSPLTEADNLS 182

QY 193 QELDTVLVNLVPAHKNLNKOAYDLAKALLKRTAQALEPYITTFNOVLGK--TSISD 250  
 DB 183 VELDULLINIVEPYKNNKFAQLTEQLTKTGDALESTIKMFFNRALVADPNTKLS- 241  
 QY 251 LSEHFDLILELYNIDSHLLSVLPQLEFKLSNDNERLQVVKVLLAKMFGAKDSELASQ 310  
 DB 242 ITNKIYDIIVELNRINAGLLCSVLPQLENKLLSTDDAERLKAITLLSRMFSEKDSOLAKK 301  
 QY 311 NKPLMOCYLGKRFNDIIVPRLCEVKFASHCLMHPDLAKDLTEYLKVRSHDPEAIRHVD 370  
 DB 302 YPNLLKIFFGRCFDITEPVRIKCVQSSMHFLLNHPSLQHDITEKRLNRHDLDEVREH 361  
 QY 371 IVSIVTAAKDKILLVND--HLLNFVRETLDKRMVRKKEAMGLAQIYKYKVALQSAAGD 428  
 DB 362 VMAIVETAKRDFTLVLEAPDLLLEIVRETLDDKKYIKRRDAMGLAIYKRAICEPNLST 421  
 QY 429 AAK-QIAWIKDKLLHIYQNSIDRLIVERIFAQYVWPHNLETTMCKLYLYIATLDLN 487  
 DB 422 GLKVRVDWIKNKLHGYYKVGLEDRLVERLLITCLVPYKLAPEERMKKLYHLGLDLDAN 481  
 QY 488 AVKALNMKQCNLLRHQVKKDLDLIKQK-TDASVKAIFSKVMVITRNLPDPGKAQDFM 546  
 DB 482 ATKAFVLOKQNMKTRNTVSDWIKLHHSKEFTPRVLSQLSAKQANAKLLPDLKAAEYL 541  
 QY 547 KKTQVLEDEDEKIRKOLEVLVSPSCCKQAEKGVRETTKKLGNPKOPTNPFLEMIKFLLE 606  
 DB 542 TQFSNNLRKDAQLLRCINIVIKRDVSCRECAWTMGVLLKGLGHAHV-SNIYNTVMKMLIE 600  
 QY 607 RIAPVHIDTESIALIKQVKNKSIDGTADDEGVTPTDQAIRAGLELLKLVLSFTHPIFSHS 666  
 DB 601 RVASVMVDKESIGVLISLIEQCIKSGMCEIGISAQAGEGRGLKLLAMLISYVFSAHFFT 660  
 QY 667 AETFESSLACLDKDEKVAEAAQIFKNTG---SKIEDPPIHRSALLPVLVHHSKKGPP 723  
 DB 661 DTSRLHLISLSLEQDYVAPVLKLTHTLGRYQPLIDDPPTAILDELAPYCKDFALIGTP 720  
 QY 724 RQAKYATHCHT-----AIFSSKTEQ-----FAQIFELHSLDPSNLEHL 763  
 DB 721 KQAKHVRCLIFVNSQSSASTDGTATSGAGSASTTQIVHPFNEIETRLKLRP-NCEHQ 779  
 QY 764 ITPLVTIGHIALLAPQDFAAPKSWATFVTKDILLMDRLPGKTKTL-----WVPDEEV 818  
 DB 780 RTKIVTLGHIAYNMPOAFLTPIKNMITARRIVKELLIOE-VPAQRDYELPEDSDWCAQELK 838  
 QY 819 SPETMVKIQAKMVMWMLCKNNHKSSTGLTLLTTLHSDGLTEQKISKPDMSRL 878  
 DB 839 PPDTLCKLDALKAMARWLLGLRTD-EHAAQKTFRMLAAAFVNRQDGLLQNRLLCAEKSRL 897  
 QY 879 RLAAAGSAIVKLAQPCPYHEITILEQYQCALAINDECYQVQVFAOKLHGLSLRLR--- 934  
 DB 898 RLGAACMLKVCQKGVGDQYSAEQYLQSLMDADVPVEIRFARKLHGLSLRSLPNC 957  
 QY 935 LPLEYMAICALCARKPKERRAHARQCLVKNINVRREYLKQHAASV-----EKLLSLP 988  
 DB 958 LPDLFMGLYVLGLETERKQLQDLVRHVAETDVNKRREYLTAVMTSPDSSTESQSLHLP 1017  
 QY 989 EYVVPYTIHLAHDPPYVKVQDIEQLKDVKECLWFVLEILMAKNENNENSHAFIRKMWENIK 1048  
 DB 1018 DYMLFAIPVLVDPRFTNHEDYVQLRMEKCLRFLEPLMAKRETFVHSFYKQLQLLIK 1077  
 QY 1049 QTKDAQCPDDAKMNEKLYTVCDVAMNIMSK-----STTYSLESPKDPVLPARFFTOP 1101  
 DB 1078 HRFSIG-SKRONYKMWALCDLAMYIIDSFPFGDGTSTFSMPL-----ALPEMYKEP 1132  
 QY 1102 D-KNFSNTKNYLPPEKMSFFTPGKPTNTVLGAVNKPLSSAGSQSQTKS 1149  
 DB 1133 AVANFQNDVYIPLDVYT-----LGA-KTSKAAATAMTTS 1167  
 RESULT 6  
 Q9HFF5  
 ID Q9HFF5 PRELIMINARY; PRT; 1205 AA.







```
QY 421 LQSAAGKDAKQIAWIKDKLLHYIYQNSIDRRLLVERIFAQYVMPHNLETTERMKCLYYL 480
Db 421 LQSAAGKDAKQIAWIKDKLLHYIYQNSIDRRLLVERIFAQYVMPHNLETTERMKCLYYL 480
QY 481 YATLDLNAVKNEMWKCQNLLRHQVKDLDLQKPKTDASVKAIFSKVWVITRNLPDPG 540
Db 481 YATLDLNAVKNEMWKCQNLLRHQVKDLDLQKPKTDASVKAIFSKVWVITRNLPDPG 540
QY 541 KAQDFMKKFTQVLEDEKTRKOLEVLVSPSCSKQAEGCVREITKGLGNPKOPTNPFLEM 600
Db 541 KAQDFMKKFTQVLEDEKTRKOLEVLVSPSCSKQAEGCVREITKGLGNPKOPTNPFLEM 600
QY 601 IKFLLERIAPVHIDTESISALIKQVKNKSIDGTADDEGVPTDQAIRAGLELLKVLFSFH 660
Db 601 IKFLLERIAPVHIDTESISALIKQVKNKSIDGTADDEGVPTDQAIRAGLELLKVLFSFH 660
QY 661 PISFHAETTESLACLMKDDKVAEALQIFKNTGSKTEEDPPIRSALLPVLHHSKK 720
Db 661 PISFHAETTESLACLMKDDKVAEALQIFKNTGSKTEEDPPIRSALLPVLHHSKK 720
QY 721 GPPRAKYAIHCHTAFSSKETQFAIFELPLHKSIDPSNLEHLITPLVTIGHIALLAPDQ 780
Db 721 GPPRAKYAIHCHTAFSSKETQFAIFELPLHKSIDPSNLEHLITPLVTIGHIALLAPDQ 780
QY 781 FAAPKSWVATFIVKDLLMNDRLPGKTTKLWVPDEVSPETWVKIQAIKMVMRWLLGKM 840
Db 781 FAAPKSWVATFIVKDLLMNDRLPGKTTKLWVPDEVSPETWVKIQAIKMVMRWLLGKM 840
QY 841 NNHSGSTLRLTLILHSDGDLTQGGIKSKPDMRSLRLAGASIVKIAQEPICYHEIT 900
Db 841 NNHSGSTLRLTLILHSDGDLTQGGIKSKPDMRSLRLAGASIVKIAQEPICYHEIT 900
QY 901 LEQYOLCALAINDECYQVQVQFAQLHGLSLRLPLEYMAICALCAKDPVKERRAHQ 960
Db 901 LEQYOLCALAINDECYQVQVQFAQLHGLSLRLPLEYMAICALCAKDPVKERRAHQ 960
QY 961 CLVKINVRREYLKQHAAYSEKLLSLLPEYVPTTHLLAHPDYVVKQDIEQLKDVKEC 1020
Db 961 CLVKINVRREYLKQHAAYSEKLLSLLPEYVPTTHLLAHPDYVVKQDIEQLKDVKEC 1020
QY 1021 LMFVLEILMAKNENSHAFIRKMWENIKOTKDAQGPDADAKMEKLYTVCDAVAMNIMSKS 1080
Db 1021 LMFVLEILMAKNENSHAFIRKMWENIKOTKDAQGPDADAKMEKLYTVCDAVAMNIMSKS 1080
QY 1081 TTYLESPPKDPVLPARFFTPQDKNFNTKNYLPPEMKSFPTPKPKTTNVLGAVNKPSS 1140
Db 1081 TTYLESPPKDPVLPARFFTPQDKNFNTKNYLPPEMKSFPTPKPKTTNVLGAVNKPSS 1140
QY 1141 AGKQSTKSRMETVSNASSNPSPGRIKGRDSEMDHSENEYTWSSPLPGKKSQ 1200
Db 1141 AGKQSTKSRMETVSNASSNPSPGRIKGRDSEMDHSENEYTWSSPLPGKKSQ 1200
QY 1201 RQSDLVRLSELEKPRGRKTPVTEBEKLGMDDLTKLVQEQPKGQSRKRKRGHTASESD 1260
Db 1201 RQSDLVRLSELEKPRGRKTPVTEBEKLGMDDLTKLVQEQPKGQSRKRKRGHTASESD 1260
QY 1261 EQOWPEEKLEKEDILENEDEQNSPPKGRGRRPPKPLGGGTKEEPTMTKSKGSKKSG 1320
Db 1261 EQOWPEEKLEKEDILENEDEQNSPPKGRGRRPPKPLGGGTKEEPTMTKSKGSKKSG 1320
QY 1321 PPAPEEEEEEQSGNTEQKSKSHQVRRAQOQRAESPESSESAIESTQSTPOKGRGRPSK 1380
Db 1321 PPAPEEEEEEQSGNTEQKSKSHQVRRAQOQRAESPESSESAIESTQSTPOKGRGRPSK 1380
QY 1381 TPSPSPQKKNV 1391
Db 1381 TPSPSPQKKNV 1391
```

RESULT 3

Q9Y215

ID Q9Y215 PRELIMINARY; PRT: 1416 AA.

AC Q9Y215;

```
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KIAA0979 PROTEIN (FRAGMENT).
GN KIAA0979.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023196; BAA76823.1; -.
DR InterPro: IPR000637; -.
DR PRINTS: PR00929; ATHOOK.
FT NON_TER 1
SQ SEQUENCE 1416 AA; 161192 MW; 0BC682D7F4657424 CRC64;
```

```
Query Match 97.2%; Score 6994; DB 4; Length 1416;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1356; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 32 VRLKMWVKTFTMDMDQDSEEEKELYLNLAHLASDEFLLKHPGKDVRLLVACCLADIFRIY 91
Db 1 VRLKMWVKTFTMDMDQDSEEEKELYLNLAHLASDEFLLKHPGKDVRLLVACCLADIFRIY 60
QY 92 APEAPYTPDKLIDPMFITRQLKGLDTSKPOFNRYFYLLLENIAWVSYNICELEDSN 151
Db 61 APEAPYTPDKLIDPMFITRQLKGLDTSKPOFNRYFYLLLENIAWVSYNICELEDSN 120
QY 152 EIFTQLYRLTFSVINGNHQKVHMHVMDLMSIICGDTVSQELLDTVLVNLVPAHKNLN 211
Db 121 EIFTQLYRLTFSVINGNHQKVHMHVMDLMSIICGDTVSQELLDTVLVNLVPAHKNLN 180
QY 212 KQAYDLAKALLKRTAQAIPEYITTFNQVLMGLKTSISDLSEHVFDLILELYNIDSHLL 271
Db 181 KQAYDLAKALLKRTAQAIPEYITTFNQVLMGLKTSISDLSEHVFDLILELYNIDSHLL 240
QY 272 SVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWCYLGFRNDIHPRL 331
Db 241 SVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWCYLGFRNDIHPRL 300
QY 332 ECVKFASHCLMHPDLAKDLTEYLKVRSHDPEAIRHDIIVISIVTAAKDILLVNDHLLN 391
Db 301 ECVKFASHCLMHPDLAKDLTEYLKVRSHDPEAIRHDIIVISIVTAAKDILLVNDHLLN 360
QY 392 FYRERTLDKRWVRKEMGLAQIYKKYALQSAAGDAKQIAWIKDKLLHYIYQNSIDD 451
Db 361 FYRERTLDKRWVRKEMGLAQIYKKYALQSAAGDAKQIAWIKDKLLHYIYQNSIDD 420
QY 452 RLIVERIFAQYVMPHNLETTERMKCLYYLYATLDLNAVKNEMWKCQNLLRHQVKDLD 511
Db 421 RLIVERIFAQYVMPHNLETTERMKCLYYLYATLDLNAVKNEMWKCQNLLRHQVKDLD 480
QY 512 LKQPKTDASVKAIFSKVWVITRNLPDPGKAQDFMKKFTQVLEDEKIRKQLEVLSPTC 571
Db 481 LKQPKTDASVKAIFSKVWVITRNLPDPGKAQDFMKKFTQVLEDEKIRKQLEVLSPTC 540
QY 572 SKQAEGCVREITTKLGNPKOPTNPFLEMIKFLERIAPVHIDTESISALIKQVKNKSIDG 631
Db 541 SKQAEGCVREITTKLGNPKOPTNPFLEMIKFLERIAPVHIDTESISALIKQVKNKSIDG 600
QY 632 TADDEGVPTDQAIRAGLELLKVLFSFTHPISFHAETTESLACLMKDDKVAEALQI 691
Db 601 TADDEGVPTDQAIRAGLELLKVLFSFTHPISFHAETTESLACLMKDDKVAEALQI 660
```



```
DB 241 LMLGTSISDLSEHVFDTLELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
QY 301 GAKDSELASQNKPLQWCVLGRFNDIHVPTRLECVFASFASHCLMNHDPDLAKDTEYLKVRSH 360
DB 301 GAKDSELASQNKPLQWCVLGRFNDIHVPTRLECVFASFASHCLMNHDPDLAKDTEYLKVRSH 360
QY 361 DPEEAIHADVIVSVITAAKODILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYA 420
DB 361 DPEEAIHADVIVSVITAAKODILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYA 420
QY 421 LOSAAGKDAKQAIWKOKLLHIYQNSIDDRLLVERIFPAQYVMPHNLETTERMKCLYL 480
DB 421 LOSAAGKDAKQAIWKOKLLHIYQNSIDDRLLVERIFPAQYVMPHNLETTERMKCLYL 480
QY 481 YATLDNAVKALEMMWKNLRLHGVKDLDLIKOPKPTDASVKALFSPKVMVITRNLPPDG 540
DB 481 YATLDNAVKALEMMWKNLRLHGVKDLDLIKOPKPTDASVKALFSPKVMVITRNLPPDG 540
QY 541 KAQDPKKFTQVLEDEKIRKOLEVLSVPTSCCKQAEGCVREITKGLGNPKOPTNPFLEM 600
DB 541 KAQDPKKFTQVLEDEKIRKOLEVLSVPTSCCKQAEGCVREITKGLGNPKOPTNPFLEM 600
QY 601 IKFLLERAPVHIDTESISALIKQVNSIDGTADDEDEGVPTDQAIRAGLELLKVLSPFH 660
DB 601 IKFLLERAPVHIDTESISALIKQVNSIDGTADDEDEGVPTDQAIRAGLELLKVLSPFH 660
QY 661 PISFHSATFESLLACLKMDDEKVAEALQIPKNTGSKIEEDFPHRSALLPVLHKKSK 720
DB 661 PISFHSATFESLLACLKMDDEKVAEALQIPKNTGSKIEEDFPHRSALLPVLHKKSK 720
QY 721 GPPROKAYAIHCHIAFISKETQFAQIEPLHKS LDPNSLEHLITPLVTIGHIALLAPDQ 780
DB 721 GPPROKAYAIHCHIAFISKETQFAQIEPLHKS LDPNSLEHLITPLVTIGHIALLAPDQ 780
QY 781 FAAPKSWVATFIVKDLLMNDRLPGKTTKLWVPDEEVSPTMVKIQAIKMMVRLGNK 840
DB 781 FAAPKSWVATFIVKDLLMNDRLPGKTTKLWVPDEEVSPTMVKIQAIKMMVRLGNK 840
QY 841 NNHKSSTLRLTILHSDGLTEQGIKSKPDMRSRLRLAAGSAIVKLAOPCYHEIIT 900
DB 841 NNHKSSTLRLTILHSDGLTEQGIKSKPDMRSRLRLAAGSAIVKLAOPCYHEIIT 900
QY 901 LEQYOLCALAINDECYQVRQVPAQKHLKGLSRLRPLEYMAICALCAKDPVKERRAHARQ 960
DB 901 LEQYOLCALAINDECYQVRQVPAQKHLKGLSRLRPLEYMAICALCAKDPVKERRAHARQ 960
QY 961 CLVKNINVRREYLKQHAASVEKLLSLLPEYVVPYTIHLLAHDPPDYKVDIEQLKDVREC 1020
DB 961 CLVKNINVRREYLKQHAASVEKLLSLLPEYVVPYTIHLLAHDPPDYKVDIEQLKDVREC 1020
QY 1021 LWFVLEILMAKNENSHAFIRKMWENIKQTKDAQGPDPAKMKNEKLYTVCDAVMNIIMSKS 1080
DB 1021 LWFVLEILMAKNENSHAFIRKMWENIKQTKDAQGPDPAKMKNEKLYTVCDAVMNIIMSKS 1080
QY 1081 TTYLESKDPVLPARFTQPKDNFSNTKNYLPPEMKSFPTPGKPTTNVLCVAVNKP LSS 1140
DB 1081 TTYLESKDPVLPARFTQPKDNFSNTKNYLPPEMKSFPTPGKPTTNVLCVAVNKP LSS 1140
QY 1141 AKQSQTSSRMETVSNASSSNPSPGRIGRLDSSSEMDHSENEDEYTWSSPLPKKSDK 1200
DB 1141 AKQSQTSSRMETVSNASSSNPSPGRIGRLDSSSEMDHSENEDEYTWSSPLPKKSDK 1200
QY 1201 RDDSPLVSELEKPRGRKKTPTVEQEELGMDLTKLVQEQPKGSQRKRGRHGTASESD 1260
DB 1201 RDDSPLVSELEKPRGRKKTPTVEQEELGMDLTKLVQEQPKGSQRKRGRHGTASESD 1260
QY 1261 EQQWPEERLKEIDILENEDEQNSPPKGRGRPPKPLGGGTPEEPTMTKTSKSGSKSG 1320
DB 1261 EQQWPEERLKEIDILENEDEQNSPPKGRGRPPKPLGGGTPEEPTMTKTSKSGSKSG 1320
QY 1321 PPAPPEEEEEERQSGNTQKSKSQHRRVSRRAQRAESPSSAISTOSTPQKGRGRPSK 1380
DB 1321 PPAPPEEEEEERQSGNTQKSKSQHRRVSRRAQRAESPSSAISTOSTPQKGRGRPSK 1380
```

```
DB 1321 PPAPPEEEEEERQSGNTQKSKSQHRRVSRRAQRAESPSSAISTOSTPQKGRGRPSK 1380
QY 1381 TPSPSQPKNV 1391
DB 1381 TPSPSQPKNV 1391
RESULT 2
QY451 PRELIMINARY; PRT: 1391 AA.
AC QY451:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN.
GN AS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
RT "Androgen-induced proliferative shutoff in prostate cancer cells.";
RL Proc. Am. Assoc. Cancer Res. 37:223-223(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=92229875; PubMed=10215036;
RA Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
RT "Early gene expression during androgen-induced inhibition of
RT proliferation of prostate cancer cells: a new suppressor candidate on
RT chromosome 13, in the BRCA2-Rb1 locus.";
RL J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
DR EMBL; U95825; AAD22134.2; -.
DR InterPro; IPR002687; -.
DR ProDom; PD004104; -.
SQ SEQUENCE 1391 AA; 158035 MW; F58AEE5AD6D9479 CRC64;
```

```
Query Match 98.7%; Score 7101; DB 4; Length 1391;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1377; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 1 MAHSTRNDGKITVPGVKEISKSEEMVRRLLKMWVKTMDMDQDSEEEKELYLNLA 60
DB 1 MAHSTRNDGKITVPGVKEISKSEEMVRRLLKMWVKTMDMDQDSEEEKELYLNLA 60
QY 61 LHLASDFELKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLDFIMFITRLQKLEDT 120
DB 61 LHLASDFELKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLDFIMFITRLQKLEDT 120
QY 121 KSPQNRIFYLLENATWKNYNICELEDSNEIFQLYRTLFESVINGHKNQVHHMVDL 180
DB 121 KSPQNRIFYLLENATWKNYNICELEDSNEIFQLYRTLFESVINGHKNQVHHMVDL 180
QY 181 MSSICEGDTVSQELLDVTVNLVPAHKNLKOAYDLAKALKRTAQAEPIYITFFNOV 240
DB 181 MSSICEGDTVSQELLDVTVNLVPAHKNLKOAYDLAKALKRTAQAEPIYITFFNOV 240
QY 241 LMLGTSISDLSEHVFDTLELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
DB 241 LMLGTSISDLSEHVFDTLELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
QY 301 GAKDSELASQNKPLQWCVLGRFNDIHVPTRLECVFASFASHCLMNHDPDLAKDTEYLKVRSH 360
DB 301 GAKDSELASQNKPLQWCVLGRFNDIHVPTRLECVFASFASHCLMNHDPDLAKDTEYLKVRSH 360
QY 361 DPEEAIHADVIVSVITAAKODILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYA 420
DB 361 DPEEAIHADVIVSVITAAKODILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYA 420
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:02:24 ; Search time 39.87 Seconds  
(without alignments)  
4615.909 Million cell updates

US-09-512-581-2  
Title: 7193  
Perfect score: 1  
Sequence: 1 MAHSGKTRTNDGKITYPGVK.....QKGRGRPSKTPSPSPQPKNV 1391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters:

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
SPREMBL_16:*
1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*
```

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	7155	99.5	1447	4	Q9NTT5	Q9ntt5 homo sapien	
2	7101	98.7	1391	4	Q9Y451	Q9y451 homo sapien	
3	6994	97.2	1416	4	Q9Y215	Q9y215 homo sapien	
4	2698.5	37.5	851	4	Q9Y4D4	Q9y4d4 homo sapien	
5	1982.5	27.6	1218	5	Q9V646	Q9v646 drosophila	
6	833.5	11.6	1205	3	Q9HF55	Q9hff5 schizosacch	
7	808	11.2	1638	10	Q9FTL0	Q9ftl0 arabidopsis	
8	807	11.2	1579	5	Q9XTF6	Q9xtf6 caenorhabdi	
9	747.5	10.4	1596	3	Q9UYV6	Q9uyv6 sordaria ma	
10	656	9.1	125	4	Q9H5N8	Q9h5n8 homo sapien	
11	632.5	8.8	1506	3	Q94076	Q94076 emericella	
12	309.5	4.3	390	3	Q94237	Q94237 schizosacch	
13	303	4.2	780	10	Q9SAI0	Q9sa10 arabidopsis	
14	264	3.7	990	10	Q9S9P0	Q9s9p0 arabidopsis	
15	248.5	3.5	2158	4	Q9Y6V3	Q9y6v3 homo sapien	
16	242.5	3.4	2265	4	Q9Y6V4	Q9y6v4 homo sapien	
17	234.5	3.3	852	10	Q9SZ55	Q9sz55 arabidopsis	
18	231.5	3.2	2253	13	P70012	P70012 xenopus lae	
19	230	3.2	2501	5	Q9NCW7	Q9ncw7 drosophila	

## ALIGNMENTS

[illegible]



**THIS PAGE BLANK (USPTO)**

---







```

1054 .....GlnGlyPro.....AspAspAlaLysMetAsnGlu 1063
1809 TCCTGCAGCAGCAGGAGGAGGCTCTAGAGCAGCGGAAGAGGCTGCAG 1858
1064 LysLeuTyrThrValCysAspValAlaMetAsnIle.....IleMetSe 1078
1859 GCAGACTGCGAGCTCTGCTGGAAGAAATGCGCTTCAAGTCCAGTC 1908
1078 rlySerThrThrTyrSerLeuGluSer..... 1087
1909 CCCTTCTATAAAATGAGCCTTGAGTCTACGGGAAGAGCTATGCTAGCA 1958
1088 .....ProLysAspProValLeuProAlaArgPhePheThrGlnProAsp 1102
1959 TGGTCCCGCAGCAATGAGAAGCTGTCGAAGACTTATGATACCACTGTGGAT 2008
1103 LysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSer... 1118
2009 GACAATGAGAGCTATTACAGAGTTTACACCAGCACCCAGACGAGCA 2058
1119 .PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValA 1135
2059 GAGCTTCTCAAGAGCTATGACAGCAGCACCACTGCCAGTGGGCTATG 2108
1135 snLysProLeuSerSerAlaGly.....LysGlnSer 1145
2109 GGAAGAGTTACTCCACTACCAGCAACAGCAGCATTTACCTATATAAGAAGCT 2158
1146 GlnThrLysSerSerArgMetLutThrValSerAsnAlaSerSerSe 1162
2159 TAGCGCAGCACCAGTAGCTCTCACACCTGCCAGAAGAGTTTGTTCAGCAG 2208
1162 r.....AsnProSerSerPro..... 1167
2209 CTGCAGTACGAGGAGAACCTGCTGAGCCTCAACAGCAGTGGAGCCCTTGAGG 2258
1168 .....GlyArgIleLysGlyArgLeu 1174
2259 AAATGGTTGTGAAGTGTGTATCAAGCTGCAGCGGTGCAGGCCATGTAC 2308
1175 aspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSe 1191
2309 CAGATAGCCAGGAGGAGACACAGCCAGCTGCAAGAG.....CAGATGGA 2352
1191 rProLeuProGlyLysLysSerAspLysArgAspSerAspLeuValA 1208
2353 AAAGTTACTGGCCACAGCAGAAAGACCTGAAGGAGAGAGCTGGATGCTGTG 2402
1208 rSerGlu.....LeuGluLysProArgGly... 1216
2403 AAAGGAGGTTCAAGGAGTGCAATGCGCTTGAAGGCCCATGGCCCCC 2452
1217 ...ArgLysLysThrProValThrGluGlnGluLysLeu..... 1229
2453 CAGAAGCAGCAAGATGAGATCAAAAGACTGCACCAAGCTCGGGAGCT 2502
1230 .....GlyMetAspLeuThrLysLeu.....V 1238
2503 GCAGCTGCAATACCAAGCTAGCATGGATGATGAGCAGGCGGCTTCTGGTAG 2552
1238 aGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHis 1254
2553 TGCAGGAGCAG.....CTGGAGGGGCAG 2575
1255 ThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArg..... 1269
2576 CTGCAGTGTCCAGGAGGAGCTCCCGCAGCTCAGGGAGAGAGAGGCTTC 2625
1270 ...LeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProp 1285
2626 TGTTCGTAAGAAAGCCCGGGGGAAGAAATCTTAATAAGAACATGAACAAGA 2675
1285 roLysLysGly...LysArgGlyArgProProLysProLeuGlyGlyGly 1300
2676 ATGCCAATGGGTTAAATGAAAAAGTGACCAAGCCATCGTCGGATACT 2725
1301 ThrProLysGluGluProThrMetLysThrSerLys..... 1312
2726 TCTGAGACGACCTTGAGACCAGAAAGAGTCTGGAGGTAGTCTGTACTA 2775
1313 .LysGlySerLysLysSerGlyProProAlaProGluGluGluGlu 1329
2776 CAAGCCAGCCAGAGAGAAATTTAGATGGACTAGCAAAAGAGAGGAGAAAGA 2825
1329 LuGluGlu..... 1331
2826 AAGAGGAGATGGAGGAGGAAAAAGAAAGTGAAGAGAGAAAGAGGAG 2875
1332 .....ArgGlnSerGln 1335
2876 CAGTGTGGGATGAGTAGTTCTGAGCCAGCAGATCTCGAAGAGCTAA 2925
1335 yAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgA 1352
2926 ATCCACAGAAAGATCAGGAGGAAAAATGAAGAGCACAAGAGAGAGAGAGA 2975
1352 laGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln 1368
2976 AGGAAGAAGACAGTGAAGAGAGGAGGATGACGCCGACTCTTCCCTTGA 3025
1369 SerThrProGlnLysGlyArgGlyArgProSerLys.....Th 1381
3026 ATCCCGAAGAAAAATACCCCTCAGACTTCCGAGAGCAAAAGTCATC 3075
1381 rProSerProSerGlnPro 1387
3076 CCTACCCCAATCCCCC 3094

```



```

556 GACTTCAGAAATGGATTCTTAGAGCTGATCCTGAAATG..... 595
      : : : : :
499 l nAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
      : : : : :
596 ..CAGTTGTTACGGCAGCAGCTACGGGAT..... 622
      : : : : :
516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIle 532
      : : : : :
623 .....GCTGAGAGCAGATGCATGCATGAAGACAAGTGTCCAGGAAT 666
      : : : : :
532 eThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysP 549
      : : : : :
667 GTGTGTGTGAGTGTGAAGAG.....CTACAGCATC 695
      : : : : :
549 heThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluVal 565
      : : : : :
696 ATGCCAGAGGTTCAGTGAGGAGGAGCAGAGG...CGGTGTCAGAGGAGCTC 742
      : : : : :
566 LeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgL 582
      : : : : :
743 .....AAGTGTGCTCAGAAAT..... 757
      : : : : :
582 uileThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuG 599
      : : : : :
758 ..... 758
599 luMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
      : : : : :
759 AGGTGCTTCGGTTTCAGACCTCCACAGTGC.....ACCCAGAAAT 799
      : : : : :
616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh 632
      : : : : :
800 GAGGAGCTGAAGTCCAGACTCTGTACCTGCAGAAAAAATATGATACTAG 849
      : : : : :
632 rAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaG 649
      : : : : :
850 CCAGGATGAGCAGAACGAGCTCTTGAAGATGCAGCTGCAACTTCAGACTG 899
      : : : : :
649 l yLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHis 665
      : : : : :
900 AGTCCGGGAGCTCAAGTGCATGAAATCCACA...CTGTAGAAAAACCA 946
      : : : : :
666 SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspG 682
      : : : : :
947 AGTGAGAGAGAGTACTGTCCCGGCTGCAGAAAGCTGCACCTCCAGCACCA 996
      : : : : :
682 uLysValAla..... 685
997 GAACGTACATGTGAGAAAGGAAAGCTGTGGAACGCGCAGCAGCTGC 1046
      : : : : :
686 .....GluAlaAlaLeuGlnIlePheLysAsnThr 695
1047 AGGAGGAGCTGCAGTGCATGCATGAGGCAGAGCTGCAGCACCTCAGGGATACG 1096
      : : : : :
696 GlySerLysIleGluGluAsp.....PheProHisIleAr 707
      : : : : :
1097 GTGGCTCTCTTCAAGAGAGCAATGAGAGGACACAGAGCAGCGCTCA 1146
      : : : : :
707 gSerAlaLeuLeuProValLeuHisIstLysSerLysLysGlyProProA 724
      : : : : :
1147 GCTTCAGGAGATGAAGCAGCTGTACCAGGCCACGAGGAGCAGCTGGAGC 1196
      : : : : :
724 rgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
      : : : : :
1197 GGCAG..... 1201
      : : : : :
741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPr 757
      : : : : :
1202 .....AAGCACATGTATGACCAGCTGGAGCAGACCTCTCTCGT 1239
      : : : : :
757 oSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleA 774
      : : : : :
1240 CTCCAGCTGGAG..... 1252
774 l aLeuLeuAlaProAspGlnPheAlaProTrpLysSerTrpValAla 790
1252 ..... 1252
791 ThrPheIleValLysAspLeuLeuMetAsnAspArgLeuPro..... 804
      : : : : :
1253 .....CTGAAAGAGCTCAAGGCTCCACCCCATTCGGGAGGACAA 1293
      : : : : :
805 .GlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProG 821
      : : : : :
1294 AGGAAAGTGTGCTAATAAG..... 1312
      : : : : :
821 luThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837
1312 ..... 1312
838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLe 854
1313 .....TG 1314
854 uThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIle 871
      : : : : :
1315 TGACACACTGCTCTCCAGACTGCAGAAATTCAGGAAAAAGTACAGGCCA 1364
      : : : : :
871 e rLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVal 887
      : : : : :
1365 GCCAGAGGAGATGGGGCAGCTG..... 1387
888 LysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTy 904
1388 .....CAGAT 1392
904 rGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnV 921
      : : : : :
1393 GGAGCAGTGTGAGCTCTCTGGAGGATCAGAGGAGGATGCAGGAGGAGCAG. 1441
      : : : : :
921 aPheAlaGlnLysLeuHisGlyLeuSerArgLeuArgLeuProLeu 937
      : : : : :
1442 .....GCCAGCTGCAGGAAGAGCTGCACAGGCTCACACTGCCACTG 1483
      : : : : :
938 GluTyrMetAlaIleCysAla..... 944
1484 CCAAGAGTGGCCTCTTACTCAAGAGTCAGGAGCTACTCACCAGTTAGA 1533
945 ....LeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArg 960
      : : : : :
1534 AGACCTGTGTGAGCTGCAGCTGCTTACCAAGGCATGCAGGAGGACAGA 1583
      : : : : :
960 l nCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
      : : : : :
1584 AGAAGCTGATACAGAACCAAGACTGT.....GTATTAAAGAACAA 1624
      : : : : :
977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPr 993
1625 TTAGAGATCCACCAAGAGCTGCCAGCTTTCAAAGAG..... 1660
993 oTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnA 1010
1661 .....TCTCATTTCCAGGAAGTGTGTGAGAATCCCG 1691
1010 spIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGlu 1026
1692 ATGATTCCAAATTTGGTAACTCTCTCCAAATGT..... 1723
1027 l leLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMe 1043
1724 .....AATCGAAACAACCAATCCAGCTGCTCATGGAGCA 1758
1043 tValGluAsnIleLysGlnThrLysAspAla..... 1053
1759 GATGCAGCCCTGCAGGTGATGTATGACCGCGGTCCAGGAGCAGCAGGAGC 1808

```



```

1692 TTCTGAACGAGGAGAAAGATGAATCTCAGGAGCATTTCTGAAGAAGAAA 1741
1282 snSerProProLysGlyLysArgGlyArgProProLysProLeuGly 1298
1742 CTACAAAGAAGAAAGCTGTTCTAGA.....TTGTCA 1773
1299 GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySe 1315
1774 GCTGGGAAGAAAGATCAGCAGGAGAGCCAGAAACAGAAAGACAGTGT 1823
1315 rlyLysLysSerGlyProProAlaProGluGluGluGluGluGluA 1332
1824 CCGTCGCAAAATCCAGTCCACAGAGAAGATTACACAGAAGCGGTCTACAG 1873
1332 rGlnSerGlyAsnThrGluGlnLysSerLysLysLysGlnHisArgVal 1348
1874 CCAACGAAGAAGAGCTGATGATGACAGTGATACA..... 1908
1349 SerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleG 1365
1909 .....ACTCCAAAGGCGTCTCTTAAGAG 1931
1365 uSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrP 1382
1932 GAAGAAGTCTGAAGAACCTTATCAAGGCTTCCCGGCTCTCTCAAAAGTCG 1981
1382 roSerProSerGlnProLysLys 1389
1982 CATCAAAAGAGAGCCAGTAAA 2004

```

seq\_name: /SID1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT46187

seq\_documentation\_block:

ID AAT46187 standard; DNA; 3586 BP.

XX AAT46187;

XX 14-APR-1997 (first entry)

XX DNA encoding cell protein homologous to the SV40 large T antigen.

XX Cellular homologue of the SV40 large T antigen; CHLA-1;  
 KW natural killer lytic associated protein; tumour suppressor;  
 KW cancer; viral infection; therapy; ds.

XX Homo sapiens.

Key	Location/Qualifiers
exon	1..3069
	/*tag= a
intron	/codon_start= 48..50
	3070..3177
	/*tag= b
exon	3178..3586
	/*tag= c
	/note= "in-frame stop codon at 3285-3287"

XX WO9638540-A1.

XX 05-DEC-1996.

XX 30-APR-1996; 96WO-US07691.

XX 30-MAY-1995; 95US-0453701.

XX (RERE-) RES DEV FOUND.

XX Fung YK;

XX WPI; 1997-034359/03.

XX P-PSDB; AAW01535.

PT DNA encoding cellular protein homologous with SV40 large T antigen -  
 PT useful for regulating initiation of DNA replication, e.g. for  
 PT treating tumours and viral infection  
 XX Example 10; Fig 11A; 59pp; English.  
 XX A DNA sequence (AAT46187) encodes a cellular homologue (AAW01535)  
 CC of the SV40 large T antigen (CHLA), characterised as a regulator of  
 CC initiation of DNA replication in cells that is useful e.g. in  
 CC treatment of tumours and inhibition of viral genome replication in  
 CC infected cells. The CHLA-1 gene was isolated from human cDNA  
 CC libraries in lambda gt10 using an antibody raised against the  
 CC retinoblastoma protein p16 motif (see also AAW01534). Plasmids  
 CC carrying the CHLA-1 gene, or a portion of it encoding amino acids  
 CC 1-146 of CHLA, can be used to transfect host cells as a means  
 CC of regulating the initiation of DNA replication in the cells.  
 CC (N.B. another CHLA-1 gene, with different 5' and 3' untranslated  
 CC region sequences, is given in AAT46186).  
 XX  
 SQ Sequence 3586 BP; 1056 A; 826 C; 1052 G; 652 T; 0 other;

alignment\_scores:

Quality:	176.00	Length:	1223
Ratio:	0.337	Gaps:	46
Percent Similarity:	42.682	Percent Identity:	17.253

alignment\_block:

US-09-512-581-2 x AAT46187 ..

Align seg 1/1 to: AAT46187 from: 1 to: 3586

```

349 LysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAl 365
62 CGTCTCTACGGGAGGAGATTCCCTGTTAGACATGAGAAGAAGCGCA 111
365 aileArg.....HisAspValIleValS 373
112 ACTTAAGAAATAGAACGAATTCATTGTGCCAGGCTGAGATCCAGA 161
373 erIleValThrAlaA.....LysLysAspIle 382
162 GTCTCGGCAACGACGAGAGATTCCGCAACTGACATGAGAGTGACATA 211
383 LeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAs 399
212 GCATCCCTGCAGGAGATCTCTGCCGGATGCAGAAATGAA.....CTGA 255
399 pLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleT 416
256 AGACATGGAAACGCATTCGGGGAGATTATGAGATGGAGATCGCCTCC 305
416 yrLysLysTyrAlaLeuGlnSerAla.....AlaGlyLysAspAlaAla 430
306 GTGCAGAAATGGAATGAAGAGCTCTGAACCATCCGTTAGTTAGTCTC 355
431 LysGlnIleAlaTrpIleLysAspLysLeuLeuHisIle.....Ty 444
356 TCAGATTACTCTGGGTTACAAGAAGAACTGCAGGAGCTCGGGAACGCTA 405
444 rTyrGlnAsnSerIleAspAspArgLeuLeuValGlu.....ArgIleP 459
406 CCATTTTCCTGAATGAGGAATACCGGGCCCTGCAGGAGACACAGCAGC 455
459 heAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
456 TCACGGGCGAGCTTCAGATCTGGAGAGTGAGAGGACACAGAGCAACA 505
476 CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLe 492
506 GAGAGATGGCTGCAGTCCCAACACTGATGATGACGTGACGAGAGTCTCA 555
492 uAsn.....GluMetTrpLysCysG 499

```







```
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154778.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 176.00    Length: 1025
    Ratio: 0.424      Gaps: 44
    Percent Similarity: 40.488    Percent Identity: 19.220

alignment_block:
US-09-512-581-2 x AAC47245
```

```
Align seg 1/1 to: AAC47245 from: 1 to: 2638
395 GluArgThrLeuAspLysArgTrpArgValArgLys.....GluAlaMe 409
    ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
28 GAGCCAAACCGCAACAGACTACTTCTTGAAGCCCATCAGAGCCTAT 77
409 tMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaG 426
    | ||| :||| :||| :||| :||| :||| :||| :||| :|||
78 GGCTGGA.....AAGGAGAATGCTGGGG 100
426 LYLysAspAla.....AlaLysGlnIleAlaTrpIleLysAspLys 439
    |||||:||||:||||:||||:||||:||||:||||:||||:||||
101 GTAAGAAACACCAAGACTGGCGGAAGATGAGGATATGCTGAGCCAGAC 150
440 LeuLeuHisIleTyrTyrClnAsnSerIleAspAspArgLeuValG1 456
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
151 AATATGGAGATAGATGCTCAGATTAAAGAAAGATGATGAA...AAGCTGA 197
456 uArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrGluA 473
    | :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 GACGGAAGATAAAGAGTCAGAGGTTAAGAAAAAATGAAGACAATGCTGAGA 247
473 tGMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVal 489
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
248 CTCAAAAAATGGAAGAGAGAGTGTGAGTCACCAAGATGAGGAGCAAGCA 297
490 LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnVa 506
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
298 GAGGCTACCAACATGATGATCAAGATCCGATGCA..... 330
506 LysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValL 523
331 .....AAGAAAGAGCAAACTGATGATGTTGTT. 357
523 ysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspPro 539
358 .....TCAGTGGAAAGATACT 372
540 GlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAs 556
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
373 GTAATGAAGGAAACGTGGAATCTAAAGACAATAACTATGCCAAAGATGA 422
556 pGluLysIleArgLysGlnLeuGluVal..LeuValSerProThrCysSer 572
    |||||:||||:||||:||||:||||:||||:||||:||||:||||
423 TGAAAAAGAGACCAAGAGACAGATATTACTG..... 454
573 CysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLysGlyAs 589
    |||||:||||:||||:||||:||||:||||:||||:||||:||||
455 .....AGGCAGACCACAAAAAAGCTGGGA... 478
589 nProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuG 606
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
479 .....AGGAGGACATACAAC.....ATGAAGCTGACAAAGCAATGGAAC 518
606 luArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIle 622
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
519 AAAGATGGCA.....ATA 532
623 LysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluG1 639
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
533 CAGGAGACATCAAGAG....GAAGGACACTGGTAGATGAAGACAAAGAG 578
639 yValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysValL 656
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
579 CACAGATATGGATGAAAAAAGTGGAGAATGGGATCAAAAAATAACAAGTGG 628
656 euSerPhe.....ThrHis 660
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
629 AGAATGTTGAAGAAAAAGAAAGAGATAGGAAGAAAAATAAACAAAG 678
661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLe 677
    :|||:||||:||||:||||:||||:||||:||||:||||:||||
679 GAAGTTGAGGGCGCAAGGCTGAGGTGATGAGTCA..... 714
```



```

6726 ...AGTTGCCAATCCCGAG.....:::|||||
1382 roSerProSerGlnPro 1387
6760 CCAGCCAGGCCCGAGCCA 6776

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA47245
seq_documentation_block:
ID_AA47245 standard; DNA; 2638 BP.
AA47245;
18-OCT-2000 (first entry)
Arabidopsis thaliana DNA fragment SEQ ID NO: 53101.
Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125786.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
04-MAY-1999; 99US-0132407.
05-MAY-1999; 99US-0132484.
06-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
21-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
```



[illegible]



```
3792  ....TATCTAGG.....GGGAAGGAATT 3811
|||||
333  sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
|||||
3812  AATGGCAGCATCTCAGGCATTTCATCTTAACCAACAGCTGGAGCAGCT 3861
|||||
350  splLeuThrGluTyr.....LeuLysVal 357
|||||
3862  CTGTAGGCCCCCACCATTAGAGAGCAACTGACCAAGTTCTACGACAGATG 3911
|||||
358  ArgSerHisAspProGluGluAlaLys.....ArgHisAspValLysVa 372
|||||
3912  CCCTCTCGAGACGACAGCACCTCGCTGACTGCCAGAGAGGAGGCCAGCAT 3961
|||||
372  lSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisL 389
|||||
3962  ACCCCGGTCTACATTAGGAGACTCAGACACAGTTGCGAGGCTGGAGAAAG 4011
|||||
389  euLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArg 405
|||||
4012  AACTGAGCAATGCCAGGAGGAG....CTGAGCTCATGCCCAAAAGAA 4058
|||||
406  LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuG 422
|||||
4059  AGAGAAAGCCAGATAGATTGCTGCGCTGCAGTCCATGATGCGCTGCA 4108
|||||
422  n.....SerAlaAlaGlyLysAspAlaAlaLysGlnI 433
|||||
4109  AGAGGAAGAGCTGCAGGTCGAGGCTGCTGACTTGGAGTCCCTGACCAGCA 4158
|||||
433  leAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
|||||
4159  ACATACAGATAAAGAGACCTCATAAAGACCTGCAAAATGCAACTGGTT 4208
|||||
450  Asp.....AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVa 464
|||||
4209  GACCTCTGAAGATATGCCAGCATGGAGCGCTGACCCCAAGAGGCTTACT 4258
|||||
464  l.....ProHisAsnLeuGluThrThrG 472
|||||
4259  TCTTCGGGAAAAAGTTGCTTCACTGGAAACCCAGGTCAGGAAGGTCAG 4308
|||||
472  lu.....ArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
|||||
4309  AGACAGGAGACAAACAGTCTCTGCTGATGTAGAGGACTAGTGGATGAA 4358
|||||
487  AsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuAr 503
|||||
4359  CGGAGT.....CGGCTCAACGAGGCCCTGCAAGCTGAG.....CG 4393
|||||
503  gHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspA 520
|||||
4394  GCAGCTCTACAGAGCGCTGGTCAAGTTCCATGCCCAACCAAGAGATCTCG 4443
|||||
520  laSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
|||||
4444  AGAGACACCGAACTCTG..... 4460
|||||
537  ProAspProGlyLysAlaGlnAspPheMetLysPheThrGlnVal.. 552
|||||
4461  .....CAGGTGGA 4468
|||||
553  LeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerP 569
|||||
4469  ACTGGAAGGGCCGAGGTGTACGAGCTGCAGTAGAA..... 4505
|||||
569  roThrCysSerCysLysGlnAlaGluCysValArgGluIleThrLys 585
|||||
4505  ..... 4505
|||||
586  LysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLy 602
|||||
```



antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;  
 antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;  
 atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;  
 eosinophilic granuloma; proliferative skin disease; ulcerative colitis;  
 reperfusion injury; atopic dermatitis; diabetes insipidus;  
 conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;  
 arterial restenosis; ankylosing spondylitis; transplant rejection;  
 graft versus host disease; ss.

Rattus sp.

Key Location/Qualifiers  
 CDS 1746..8726  
 /\*tag= a  
 /product= "Myomegalin"  
 /note= "this region is referred to as SEQ ID NO: 3  
 in the specification"

W0200027861-AL.

18-MAY-2000.

12-NOV-1999; 99WO-US26860.

12-NOV-1998; 98US-0108255.

(STRD ) UNIV LELAND STANFORD JUNIOR.

Conti M, Pahlke G;

WPI; 2000-376479/32.

P-PSDB; AAY71158.

Polynucleotide encoding a phosphodiesterase (PDE) interacting  
 polypeptide, useful for diagnosis and treatment of asthma, cystic  
 fibrosis, Crohn's disease, and rheumatoid arthritis -

Disclosure; Fig 2; 77pp; English.

The present sequence is a cDNA encoding a phosphodiesterase (PDE)  
 interacting protein, myomegalin from rat skeletal muscle cDNA library.  
 The protein modulates the functions and properties of PDEs, specifically  
 CAMP-PDEs, and also targets them to specific subcellular compartments.  
 Rat myomegalin has at least four differently sized transcripts, two of  
 which are expressed in heart (7.5 and 5.9 kb), two in skeletal muscle  
 (7.5 and 4.3 kb) and one in testis (2.5 kb). The present sequence  
 can be used in the diagnosis and treatment of disease conditions  
 associated with PDE activity. The diseases include asthma, cystic  
 fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic  
 granuloma, psoriasis, proliferative skin diseases, endotoxin shock,  
 septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,  
 inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory  
 distress syndrome, diabetes insipidus, allergic rhinitis, allergic  
 conjunctivitis, vernal conjunctivitis, arterial restenosis,  
 atherosclerosis, inflammatory diseases associated with irritation and  
 pain, rheumatoid arthritis, ankylosing spondylitis, transplant  
 rejection and graft versus host disease, disease conditions associated  
 with hypersecretion of gastric acid, and disease conditions in which  
 cytokines are mediators.

Sequence 9679 BP; 2549 A; 2626 C; 2636 G; 1867 T; 1 other;

alignment\_scores:  
 Quality: 177.00 Length: 1539  
 Ratio: 0.270 Gaps: 72  
 Percent Similarity: 42.625 Percent Identity: 18.583

alignment\_block:  
 us-09-512-581-2 x AAD00768 ..

Align seg 1/1 to: AAD00768 from: 1 to: 9679

22 lleserAspLysIleSer.....LysGluGluMetValArgAr 34  
 |||:|||||:||||| |||:|||||: |||:|||||: |||:|||||:  
 3219 ATAGATGAAAGTCTCTGCTCTACAGAAAGGACAAAGGAAGTCTCGGCA 3268  
 34 gleuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluG 51  
 |||:|||||:||||| ||| |||:|||||: |||: |||: |||: |||:  
 3269 GCTCGGGTCTGCTGTGAGG.....GACCGAGACCATGACTTAGAGA 3309  
 51 luGluLys...GluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66  
 :: :: :: ||| ||| ||| :: :: ::  
 3310 GACTCGGTGTCTCTCTGCTGCCAATGAAGCTACCATGCAAGATGAGAG 3359  
 67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCy 83  
 ::|||:|||||: ||| ::|||:|||||: |||: |||: |||: |||:  
 3360 AGTCCTCTGAGGCCAGAGGCTGGAAGTGGAGCAGATTAATTCACCACTG 3409  
 83 sleuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerP 100  
 |  
 3410 C..... 3410  
 100 roAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116  
 3410 ..... 3410  
 117 LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGl 133  
 3411 .....CA 3412  
 133 uAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspS 150  
 :|||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3413 AAACCTCCAGTGGTTGAAGGAA.....GAATTGGAA... 3443  
 150 erAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166  
 ||| ||| ||| ACCAGTTT..... 3452  
 3444 ..... 3452  
 167 AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSe 183  
 ||||| ||||| ::|||:|||||: |||: |||: |||: |||: |||:  
 3453 ...GGCCACTGGCAGAAGACACAGGAGAGCATCATTCAGCAGTACAGAC 3499  
 183 rIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 200  
 |||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3500 ATCTCTGATGACAGACGACAAAGAGTAGAGGATCTCAGTGCACACTTGC 3549  
 200 euValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAsp 216  
 |||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3550 TCCACAAACTTGGACCGCCGACAGAGTGAAGTA.....GCTGAGGAG 3590  
 217 LeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIl 233  
 |||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3591 CTGTGCCAGCGCTGCAGCGGAAG..... 3614  
 233 eThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerA 250  
 :|||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3615 .....GAAAGGGTCTG.....CAGGACCTTCTGAGTG 3642  
 250 sPLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsp 266  
 |||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3643 ATCGGAACAACAAGCCATGGAGCAGACGATGGAGTCCAGGGA..... 3686  
 267 SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSe 283  
 ||||| ||||| ::|||:|||||: |||: |||: |||: |||: |||:  
 3687 .....CTGCTCCAGTCGATG.....GG 3703  
 283 rAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetP 300  
 :|||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3704 CACCCGGGAACAGGAACAGGCTGTTGCAGAAAAAATGGTACAGCCT 3753  
 300 heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316  
 |||:|||||: |||: |||: |||: |||: |||: |||: |||: |||:  
 3754 TCATGGAAGAAACTCGGAATTACAGGCCCTCGGCAG..... 3791  
 317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333



```

3029 AAAAGTAATGAGTGTGCAGAGATGCTAAAA...CAAAAAGTAGAGGAGCT 3075
      ::::: |||::: ||| ::::: ::
832 tValArgTrrpLeuLeuGlyMetLysAsnHisSerLys..... 845
      ::::: |||::: ||| ::::: |||
3076 TACTAGG.....CAGAATAATGAACCAAAATTAATGATGC 3110
      ::::: |||::: ||| ::::: |||
846 .....SerGlyThrSer 849
      ::::: |||::: ||| ::::: |||
3111 AGAATTCAGGCAGATCAGAAATATAGTTTATAGAACAACTATC 3160
850 ThrLeuArgLeuLeuThrThrIleLeuHisSerAspLys..... 862
      |||::: |||::: |||::: |||
3161 CAGCAAGATGTGAAGCAGCTGAAATTCAGGCAGATGTTTAAAGATCA 3210
863 .....AspLeuThrGluGlnGlyLysIleSerLysP 873
      |||::: |||::: |||::: |||
3211 GCTACGAGCAGCAAAATGAACACTTGCATAACAGACAAAAACAGCAGG 3260
873 roAspMetSerArgLeuArg.....LeuAlaAlaGlySer 884
      ::::: |||::: |||::: |||
3261 ATTTTCAACAAATTAATATGCTAGAACAGACCTGGCGCAAAAGTCAA 3310
885 AlaIleValLysLeuAlaGlnGluProCysTyrHisGlnIleThrLe 901
      ::::: |||::: |||::: |||
3311 AATTGTGTAAGTGAATTTAAGCAAAAGTGTGACCAACAGAACATTTATCAT 3360
901 uGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys..... 915
      ::::: |||::: |||::: |||
3361 CCAGATACCAAGAAAGATTTAGAAATCTGAATTCGGAACCTGAATGCTT 3410
916 .....TyrGlnValArgGlnVal 921
3411 CCAAGAGAGACGCGCGGGGAGCAGAAAGTTACGCTACACAAAGCT 3460
922 PheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG 938
      ::::: |||::: |||::: |||
3461 CAGGTGCAAGAGTTAAATAACAGGTTTCAAAAAAGTACACAGCAATTTACA 3510
938 uTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 955
      ::::: |||::: |||::: |||
3511 CTTAAAGACCATTA.....GAGGAGCAGA 3533
955 rGAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
      ::::: |||::: |||::: |||
3534 TGACCCACAAAGATGGTCTGTTTCAGGAAGAATCTGGTAA..... 3577
972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuPr 988
      ::::: |||::: |||::: |||
3578 ...TTCAACAATCAGCAGAGGAGTTTCGGAAGAAGATGGAATAATTAAT 3624
988 oGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspT 1005
      |||::: |||::: |||::: |||
3625 GGAGTCCAAAGTCATCACTGAAAATGATATTTTCAGGCATTAGGCTTGACT 3674
1005 yrValLysValGln..... 1009
      ::::: |||::: |||::: |||
3675 TTGTGCTCTTCAACAGAAACTCTAGACCCCAAGAAATGCTAAGCTT 3724
1010 .....AspIleGlu...GlnLeuLysAspValLysG 1019
      ::::: |||::: |||::: |||
3725 TGTGAACAAACATTAAGAAGTCTGAAGACAGCTTCAACAGTATCGTGA 3774
1019 u.....CysL 1021
      ::::: |||::: |||::: |||
3775 ACAAATCAGCAGGAGGAGCAGATGGAAGCAAAATCATTAACAAAATGTC 3824
1021 eUrrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHis 1037
      ::::: |||::: |||::: |||
3825 AGAACTTGAGGATGAGCTGATGCCAGAACGCTGAG..... 3862
1038 AlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaG 1054
      |||::: |||::: |||::: |||

```

```

3863 .....GTTGAAAACCTGAAGCAAAAAATGGACCAACA 3894
1054 nGlyProAspAlaLysMetAsnGluLysLeuTyrThrVal..... 1068
      |||::: |||::: |||::: |||
3895 G.....ATCAAGAGCATGAACATCAATTAGTTTGTCTCCAGT 3932
1069 .....Cys.....Asp 1070
      |||::: |||::: |||::: |||
3933 GTGAATTCAAAAAAGACAGCAAGAAAGACTGTACCTTCAAAACCATAT 3982
1071 ValAlaMetAsnIle.....IleMetSerLys 1079
      |||::: |||::: |||::: |||
3983 TTTGAGATGACAGTGAAGAGTCCAGCACTCTCGAGAGCTGTCTCTAG 4032
1079 sSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaA 1096
      ::::: |||::: |||::: |||
4033 AAACACTGGACACTTCAACCAACCCAGATCCCTCTCTGTG.....A 4076
1096 rgPhePheThrGlnPro.....AspLysAsnPheSerAsnThr 1108
      |||::: |||::: |||::: |||
4077 GATGGACTCAAGAACACACAGCCATTGGAAGAGAAGTGGCAGCATCGGTT 4126
1109 LysAsnTyrLeuProGluMetLysSerPhePheThrProGlyLysPr 1125
      ::::: |||::: |||::: |||
4127 GTTGAACAGATACCAAGAAAGTCCAA...TTCAGCCACACAGG..... 4168
1125 oLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaG 1142
      |||::: |||::: |||::: |||
4169 .....GTCCTCACTCGAAGAAAGAGA 4187
1142 lLysGln.....SerGlnThrLysSerSerArg 1151
      ::::: |||::: |||::: |||
4188 AAGCCAGCAGTGTACTCTGAGTACTTTTCTCAGACAGCAGGAGTTA 4237
1152 MetGluThrValSerAsnAlaSerSerSerSerSerSerProGly 1168
      ::::: |||::: |||::: |||
4238 CAGATAACTTTT.....GATGACAAACCCCATTA..... 4270
1168 yArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnG 1185
      |||::: |||::: |||::: |||
4271 .....AGACTG.....TCTGAATTTGAGAAGATAAGAGACC 4301
1185 luAspTyrThrMetSerSerProLeuProGlyLysLysSerSerLysArg 1201
      ::::: |||::: |||::: |||
4302 AAGCCCTGAACAATCTAGACCA...CCTGTAGGTATCAAGATAAGCCA 4348
1202 AspAspSerAspLeuValArg.....SerGluLeuGluLysProArgG 1216
      ::::: |||::: |||::: |||
4349 TGTGAATGGAAGTGTGAAGTTTTCACACCTTTAGAGATAGTAAGAA 4398
1216 yArgLysLysThrProValThrGluGlnGluLysLeuGlyMetAspA 1233
      ::::: |||::: |||::: |||
4399 CAAGCAGTATGATATGCATACAGAA..... 4423
1233 spLeuThrLysLeuValGlnGlnLysPro 1243
      ::::: |||::: |||::: |||
4424 ..GTCACACATTTAAACAGAAAGAACCCA 4453

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAD00768

seq\_documentation\_block:

ID AAD00768 standard; cDNA; 9679 BP.

XX AAD00768;

AC AAD00768;

XX 08-SEP-2000 (first entry)

XX Rat phosphodiesterase interacting protein, myomegalin cDNA.

XX Rat; phosphodiesterase interacting protein; myomegalin; PDE; CAMP-PDE;  
 KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;  
 KW antiinflammatory; antipruritic; dermatological; shock;  
 KW analgesic; immunosuppressive; antiulcer; vasotropic; antiarthritic;



[illegible]











```

1873 ATCACTGACCTG.....CAGAGCTCCATCTC 1898
249 rAspLeuSerGlu.....Hisv 255
|:|||||:
1899 CAACTCAGCCAGCCCAAGAGAGCTGGAGAGGCTCCAGGCTCATG 1948
255 alphaSpLeuLeuGluLeuTyraSniLeAspSerHisLeu..... 269
|:|||||:
1949 GGGCCCGGTTGACTGCCAGGTGGCTCTCTGACCTCTGAGCTCACCACA 1998
270 LeuLeuSerValLeuProGlnLeuGluPheLys.....LeuLysSe 283
|:|||||:
1999 CTAATGCCACCTCCAGCAACAGACTCAAGAACTGGCTGGCTGAAGCA 2048
283 rAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetP 300
|:|||||:
2049 GCAGGCCAAGAGAACAGCCCGCTAGCAGACAGCCCTC..... 2088
300 heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
|:|||||:
2089 .....CAACAGCAAGAACAGGCTCCAGGGC..... 2115
317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333
|:|||||:
2116 .....CTCCGCCACCAGGT 2129
333 sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
|:|||||:
2130 GGAGCAGCTAAGCAGTAGCTGAAGCAGAGGAGCAGCAGTTG...AAGG 2176
350 spLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
|:|||||:
2177 AGGTACCGGAG.....AAGCAGGAGGCAACT 2202
367 ArgHisAspValIleValSerIleValThr..... 376
|:|||||:
2203 AGCAGGAGCATTGCCAGCAACTGGCCACTGCTGCAGAGGAGCAGAGGC 2252
376 ..... 376
2253 CTCCTTAAGGAGCGGGATGGCGCTCTCAAGCAGCTGGAGGCACTGGAGA 2302
377 .....AlaAlaLysLysAspIleLeuValAsnAspHisLeuLeu 390
|:|||||:
2303 AGGAGAAGGCTGCCAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGCT 2352
391 AsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGl 407
|:|||||:
2353 AATGAAGCCCGGACAGTGCACAGCTCAGTGACACAGGCCCGCCAGCGGA 2402
407 uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr...AlaLeuGlnS 423
|:|||||:
2403 GAAG.....GCAGAGCTGAGCGCGAAGGTGGAGGAACCTCCAGG 2440
423 erAlaAla.....GlyLysAspAlaAlaLysGlnIle 433
|:|||||:
2441 CTTGTCTGAGACAGCCCGCCAGCAACATGATGAGCCCGCCAGGTT 2490
434 AlaTrpIleLysAspLysLeuHisIleTyrTyrGlnAsnSerIleAs 450
|:|||||:
2491 GCAGAGCTAGTTCAGCTGCGGTCTGAGCAGCAAAAAAGCAACTGAGAA 2540
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
|:|||||:
2541 AGAAGGGTGGCCAGAGAGGAGCAGCTCCAGGAGCAGCTCCAGGCC 2590
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
|:|||||:
2591 TCAAGAGCTCTTGAAGGTCAACAAG.....GCAGC 2622
484 LeuAsp.....LeuAsnAlaValLysAlaLeuAsnGluMetTrpLys 497
|:|||||:
2623 CTTGAAGAGGAGAAAGCGGAGTGCAGATGCCCTGGAAGAGCAGCAGCG 2672

```

```

497 sCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIle 514
|:|||||:
2673 TTCTATCTCTGAGCTCAAGCAGACAGCCCGACCTGGTGGACGCATA 2722
514 ysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
|:|||||:
2723 AGCGGGAACGAAAGGAG..... 2739
531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLy 547
2739 ..... 2739
547 sLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuG 564
|:|||||:
2740 .....CTGGAAGAAGAGAGGCTGGCCGCAAGGGCTGG 2773
564 luValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
|:|||||:
2774 AGSCTCGATTACAGCAGCTTGGGAGGCCCATCAGGCTGAG..... 2814
581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPh 597
|:|||||:
2815 ACTGAAGTCTCGCG..... 2829
597 eLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleA 614
|:|||||:
2830 .....CGGAGCTGCCAGAGCCCTGGCTGCCACGACACACAG 2866
614 spThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
|:|||||:
2867 CTGAGAGTAGTGTGAGCAGCTCGTCAAGAACTAGCTGCTGCGCTGAG 2916
631 GlyThrAlaAspAspGluGlyValProThrAspGlnAlaIleAr 647
|:|||||:
2917 CGGTATGAGGATAGCCAGCAAGAGGAGCAGATATGGCCCATGTCCA 2966
647 gAlaGlyLeuLeuLysValLeuSerPheThrHisProIleSerP 664
|:|||||:
2967 GGACAGCTGATGACTTTGAAG..... 2988
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
|:|||||:
2989 .....GAGGAATGTGAGAAGCCCGCCGAGGCTGCAGGAGGCA 3027
681 AspGluLysValAla.....GluAlaAlaLeuGlnIlePheLy 693
|:|||||:
3028 AAGGAGAAGGTGCGAGCATAGAATCCACAGCGAGCTCCAGATA..... 3072
693 sAsnThrGlySerLysIleGluAspPheProHisIleArgSerAlaL 710
|:|||||:
3073 .....AGCCGCGCAGCAAGCAACTAGCTGAGCTCCATGCCCAAC 3112
710 euLeuProValLeuHisLysLysLysGlyProProArgGlnAla 726
|:|||||:
3113 TGCCAGAGCACTCCAGCAGGTCCAGAGAAG...GAAGTCAGGCGCCAG 3159
727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysLysThrGl 743
|:|||||:
3160 AAGCTTGCAGATGAC.....CTCTCCACTCTGCAGGAGAAA 3194
743 nPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnL 760
|:|||||:
3195 GATGGCTGCC.....ACCAGCAAGAGGTGGCCGCT 3226
760 euGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
|:|||||:
3227 TGGAGACCTTGGTG..... 3240
777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIl 793
|:|||||:
3241 .....CGAAGGCGAGGTGAGCAGCAGGAAACAGCTCCCGGGAGTT 3281

```







```

:|||||: 3042 AGTCAAGGAGCCTCGAGGCGAGGCGAGACACAGCCGAG. .... 3081
809 hrLysLeuTrpValProAspGluGluValSerProGluThrMetValLys 825
:|||||: 3082 .....TGGCTG.....GAAGAGCAACAGGAGCGCCCTTCTGCAGC 3117
826 ileGlnAlaIleLysMetMetValArgTrpLeuGluMetLysAsnAs 842
:|||||: 3118 ACAGCGCAGCG.....CTCAGGCTATGGAGCGGA 3149
842 nHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuH 859
:|||||: 3150 GGCAGACAGATGGCAATGAGTGAACGGCTCGCGCGCGCTGATGG 3199
859 isSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
:|||||: 3200 AGAGCCAGGCGCAGCAGCAGGAGCGTGGGAGCAGGAAGGAGGTG 3249
876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnG 892
:|||||: 3250 GCGCGGTGACCCAGAGCGGCGCGTGCAGGCTGACCTTGCCCTGGA 3299
892 uProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAla 909
3300 G..... 3300
909 euAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLys 925
3300 ..... 3300
926 LeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaI 942
3300 ..... 3300
942 eCysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaA 959
3301 .....AAGCGGCCAGAGCAGAGCTG 3322
959 rgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
3323 AGATGCGGCTGCAGAGAGCGCTCAACAGCAGCGT.....GTGGAG 3363
976 HisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrVal 992
3364 TTGCTACCTCGCAAGAGGCACCTGCTGCTGACCGCTGACGGAAAGGAA 3411
992 lProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysVal 1009
3412 .....GGCAGGACAGGAGTGGCCAGCTTC 3439
1009 inAspIleGlu.....GlnLeuLysAspValLysGluCysLeuTrpPhe 1023
:|||||: 3440 GTGCTCTGGAGGCGCCAGACAGTAAAGAGCTGGAGGAACCTCGGCAACC 3489
1024 Val.....LeuGluIleLeuMetAlaLysAsnGluAsnSerHisAl 1038
:|||||: 3490 GTAAGCAACTGAAGGAACAGCTGGCTAAGAAAGAAAGAGGACCGCATC 3539
1038 aPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnG 1055
3540 TGGCTCAGGAGCCCAACTCTGAGGCTGCTGCGCAGGACAG. .... 3579
1055 lyProAspAlaLysMetAsnGluLysLeuTyrThrValCysAspVal 1071
3580 ..CCAACAGCCCAAGCTGGGCACTGCGGCGCAGAGGTGAGCAAGCTG 3627
1072 AlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerPr 1088
3628 GACAGCAATGCCAGAGCAGGAGCAGGCTGACAGCCTGGAAACGCAG 3677
1088 oLysAspProValLeuProAlaArgPheThrGlnProAspLysAsnP 1105
:|||||:

```

```

3678 CCTCGAGGCTGAGCGGCCCTCCCGG.....GCTAGCGGGACAGTGTCTC 3721
1105 heSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThr 1121
:|||||: 3722 TGGAGACTTCGACGGGCGAGTTAGAGGAGAGGCCAGGAG..... 3762
1122 ProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
:|||||: 3763 .....CTAGGCGACAGTCAGAGTGCCTT 3785
1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
:|||||: 3786 AGCCTCGGCCCAACGGGAGTTGGCTGCTCCGCAACCAAGGTACAAGACC 3835
1155 ValSerAsnAlaSerSerSerSerSerAsnProSerSerProGly..... 1168
3836 ACAGCAAGGCTGAGATGAGTGAAGGCCAGGTCGCCCGGGCGGCCGCAA 3885
1169 ArgIleLysGlyArgLeuAspSerSerGluMetAsp..... 1180
3886 GAGCTGAGAGGAAAAATAGCTCATCAGCAGCTTGGAGGAGGAGGTGTC 3935
1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysL 1197
3936 CATCCTGAA.....TCGCCAGGTCTCGGAGAAGG 3964
1197 ySerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLys 1213
3965 AGGGGAGAGCAAGGAGTTGAAGCGGTGGTGATGCCCGAGTCAGAGAAG 4014
1214 ProArgLysLysLysThrProValThrGluGlnGluLysLeuG 1230
4015 AGCAG.....AAGCTGGAGGAGAGGCTG... 4038
1230 yMetAspAspLeuThrLysLeuValGlnGlnLysProLysGlySerG 1247
4039 .....CGCCTGTGTCAGGCGAGACAGACCCAGCAACATG 4072
1247 InArgSerArgLysArgGly.....HisThrAla 1256
4073 CCAGAGCTGCAGAACCGACGCTCTGCTGCGGGAGGAGGTGCAGAGCTC 4122
1257 SerGluSerAspGluGlnTrpProGluGluLysArgLeuLysGluAs 1273
4123 CGGAGAGGCTGAGAAACAGCGGTGCTTCAGAGAACCTGCGGCGAGGA 4172
1273 pIle.....LeuGluAsnGluAspGluGlnAsnSerProProLys. 1286
4173 GCTGACCTCACAGGCTGAGCGTGGGAGGAGCTGGGCCAAGAATTGAAGG 4222
1287 LysGlyLysArg.....GlyArgProProLysProLeuG 1298
4223 CGTGGCAGGAGAAGTTCTCCAGAAAGAGCAGCGCTCTCCACCTGCAG 4272
1298 yGlyGlyThrProLysGluPro.....ThrMetLysThrS 1311
4273 CTCAGACACACCACACAGGCCCTGGTGAGTGAGCTGCTGCCAGCTAA 4322
1311 erLysLysGlySerLysLysSerGlyProProAlaProGluGluGlu 1327
4323 GCACCTCTGCCAGCAGCTGCAGCGCAGCAGCGCTGCCGAGA..... 4366
1328 GluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLy 1344
4367 .....AACCCACCGCTGAGGAGCTGGAGCAG 4392
1344 sGlnHisArg..... 1347
4393 AGCAAGCAGCGCTGGGGAGCTGCGGGCAGAGCTGCTGGGGGCCAGCG 4442
1348 .....ValSerArgAlaGlnGlnAla 1356
4443 GGAGCTTGGGAGCTGATTCCTCTGCGGAGAGGTGGCAGAGCAGGAGC 4492

```















[illegible]

```

3553 ATTAAATGAGCA.....:|||||:
409 eMetGlyLeuAlaGlnIleTyrLysTyrAlaLeuGlnSerAlaAla 425
3585 TTTTAAATTGGAAGATGATTAAA.....:|||||:
426 GlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHI 442
3610 .....:|||||:
442 sIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIleP 459
3611 .....AGTGAAGTGCATGCTAATTACTCTAGAAGAAATTA 3644
459 heAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
3645 AGGATGAA.....CCGGTCAAAAGAG.....:|||||:
476 CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLe 492
3668 .....GTAGAAAAGAACTGTTAGTATTAT 3693
492 uAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspL 509
3694 TGAAGAAATG.....GAAGAAATATTCTAGATG 3722
509 euleuAspIleuLeuLysGlnProLysThrAspAlaSerValLysAlaIle 525
3723 TATTAGAGGAAGAAAAGATTTAAACAGACAGATGATAGATGCAGTA 3772
526 PheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
3773 GAAGAAATCATTGAATATCTTCAGATTCTAAAGAGAACTGAATCTAT 3822
542 acIlnAspPheMetLysLysPheThrGlnValLeuGluAsp.....:|||||:
3823 TAAAGATAAAGAAAAGATGTTTCTACTAGTGTGTCGAAGAAGTTCAGACA 3872
556 .....AspGluLysIleArgLysGlnLeuGluValLeuValSerPro 569
3873 ATGATATGATCAAGATGTTGAGAAGTTTGAAGATGAAAATAATGCGAA 3922
570 ThrCysSerCysLysGlnAlaGluCysValArgGluIleThrLysLys 586
3923 GAGGAGTTAATCAAGGATGCTTGCA...ATGAATGCATTACTACGAA 3969
586 sLeuGlyAsnProLysGlnProThrAsnProPhe...LeuGluMetIleL 602
3970 ACTTATTGAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATA 4019
602 yspPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIle 618
4020 AA.....GATATGAAAATAATTA 4036
619 SerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAs 635
4037 AAGAATTAGAAAAGCAATTATCAGAAGATCTTAAGAAATAATAGATGC 4086
635 pGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluL 652
4087 AAAAGATGAT.....ACAT 4100
652 euleuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGlu 668
4101 TAGAAAAGATTTATTAAAGAGGAACATGATATAAGC.....ACG 4138
669 ThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAl 685
4139 ACGTTGGATGAAGTTGTAGAATTAAGATGTCGAAGAAGCAAGATCGCA 4188
685 agIluAlaIle.....LeuG 690

```







2344 ACTGAAGTCTCTGGG..... 2358  
597 eLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleA 614  
2359 .....CGGAGCTGGCAGAGCCATGGCTGCCACACACAG 2395  
614 spThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630  
2396 CTGAGAGTGAAGTGTGAGCAGCTCTCAAGAAGTAGCTGCCTGGCTGAC 2445  
631 GlyThrAlaAspGluAspGluGlyValProThrAspGlnAlaIleAr 647  
2446 GGCTATGAGCATAGCCAGCAAGAGGACACTATGGGCCCATGTCCA 2495  
647 gAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerP 664  
2496 GGAACAGCTGATGACTTTGAAG..... 2517  
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680  
2518 .....GAGGAATGTGAGAAGCCCGCCAGAGCTGCAGAGGCA 2556  
681 AspGluLysValAla.....GluAlaAlaLeuGlnIlePheLys 693  
2557 AAGGAGAAGTGGCAGCATAGAATCCACAGCAGAGCTCCAGATAAGCCG 2606  
693 sasnThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaL 710  
2607 GCAGCAG...AACAACTAGCTGAGCTCCATGCCAACCTGCCAGAGCAC 2653  
710 euLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726  
2654 TCCAGCAGGTC..... 2664  
727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrG1 743  
2665 .....CAAGAGAAGGAAGTCAG 2681  
743 nPheAlaGlnIlePheGluProLeuHis..... 752  
2682 GGCCAGAGAGTTCAGATGACCTCTCCACTCTGCAGAAAGATGGCTG 2731  
753 ..LysSerLeuAspProSerAsnLeuHisLeuIleThrProLeuVal 768  
2732 CCACCACAAGAGTGGCCGCTTGAGACCTTGGTC..... 2769  
769 ThrIleGlyHisIleAlaLeuAlaProAspGlnPheAlaAlaProTr 785  
2770 .....CGCAAGGCAGGTGAGCA 2786  
785 pLysSerTrpValAlaThrPheIleValLysAsp...LeuLeuMetAsnA 801  
2787 GCAGGAACAGCCCTCCGGGAGTTAGTCAAGAGACCTGCGAGGCGCAGAG 2836  
801 spArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817  
2837 ACAGACAGCCCGAG.....TGGCTG.....GAAGAG 2862  
818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValAr 834  
2863 CAACAGGAGCCCGAGTCTGCAGCACACAGCAGCG..... 2898  
834 gTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrL 851  
2899 ....CTCAGGCTATGGAGGGGAGGAGCAGACAGATGGCAATGAGCTGG 2944  
851 euArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGln 867  
2945 AACGGCTGGGGCCGCTGATGAGAGCCAGGGGAGCAGCAGCAGGAGGAG 2994  
868 GlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaIleGlyse 884  
.....:.....:|||||

2995 CQTGGCAGCAGGAAAGGAGGTGGCGCGCTGACCCAGGAGGGGGCCG 3044  
884 rAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrL 901  
3045 TGGCCAGGCTGACCTTGCCTGGAG..... 3069  
901 euGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917  
3069 ..... 3069  
918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuAr 934  
3069 ..... 3069  
934 gLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProv 951  
3069 ..... 3069  
951 allLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsn 967  
3070 ..AAGCGGCCAGCAGAGCTTGAGATCGGCTGCAGAACGCCCTCAAC 3117  
968 ValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLe 984  
3118 GAGCAGCGT.....GTGGAGTTGCTTACCTGCAAGAGCACTGGC 3158  
984 uSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001  
3159 TCATGCCCTGACGGAAGGAA.....GGCA 3184  
1001 isAspProAspTyrValLysValGlnAspIleGlu.....GlnLeuLys 1015  
3185 AGACACAGGAGTTGGCCAGCTTCGTGGTCTGGAGCAGCCAGATAAAA 3234  
1016 AspValLysGluCysLeuTrpPheVal.....LeuGluIleLeuMetal 1030  
3235 GAGCTGGAGGAAGTTCGGCAACCGTGAAGCACTGAAGGAACAGCTGGC 3284  
1030 aLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnI 1047  
3285 TAAGAAAGAAAGGAGCAGCCATCTGGCTCAGGAGCCCAATCTGAGGCTG 3334  
1047 leLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGlu 1063  
3335 CTGGCAGGACAGAG.....CCAACAGGCCCCCAAGCTGGAAGCA 3372  
1064 LysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080  
3373 CTGGGGCAGAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGA 3422  
1080 rThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgp 1097  
3423 GCAGGCTGACAGCTGGAGCGACCTCGAGGCTGAGCGGGGCTCCCGG.. 3471  
1097 hePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113  
3472 .....GCTGAGCGGACAGTCTCTGGAGACTCTGCAGGCCCATTTAGAG 3516  
1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnVa 1130  
3517 GAGAAGCCCGCAGGAG..... 3531  
1130 lLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnT 1147  
3532 .CTAGGCGACAGTCAGAGTGCCTTAGCTCGGCCCAACGGGAGTTGGCTG 3580  
1147 hr..LysSerSerArgMetGluThrValSerAsnAlaSerSerSerAs 1163  
3581 CCTTCCGCCCAAGGTACACAGCACAGCAGAGGCTGAAGATGAGTGAAG 3630  
1163 nProSerSerProGly.....ArgIleLysGlyArgLeuAspSerS 1177  
3631 GCCCAGGTGGCCGGCGCGCAAGAGGCTGAGAGGAAAAAATAGCCTCAT 3680



```

55  LeuTyrLeuAsnLeu.....AlaLeuHisLeuAlaSerAspPhePh 68
||| :|:|:|:|:| ||| |
967 CTTTCTTTAAGCTGGGAGTTTGCACATCATCTG..... 1002
68  eLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuA 85
1002 ..... 1002
85  laAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAsp 101
1003 .....CAG 1005
102  LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGl 118
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1006 CAGCTACAGGATGCCCTCAATGAGCTGACGGAG.....GA 1040
118  uAspThrLysSerProGln.....PheAsnArgTyrPheTyrLeuLeuG 133
| :|:|:|:|:|:| ||| :|:|:|:|:|:| :|:|:|:|:|
1041 GCACAGCAAGGCCACTCAGGAGTGGCTAGAGAAAGCAGGCCAGCTGGAGA 1090
133  luAsnIleAlaIleValLysSerTyrAsnIleCysPheGluLeuGluAsp 149
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1091 AGGAGCTCAGCCAGCCCTGCAGGACAAGAATGC.....CTTGAAGAG 1134
150  SerAsnGluIle.....PheThrGlnLeuTyrArgThrLeuPh 162
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1135 AAGAACGAAATCTTTCAGGGAAAACTTTCACAGCTGGAAGAACACTTTC 1184
162  eSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValA 179
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1185 CCAGCTGCAGGATACCCACCCAGGAGAGAGGGCGAGGTG..... 1224
179  spLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 195
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1225 .....CTGGGTGATGCTCTTCAGCTGGAAACC 1251
196  LeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuLys 212
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1252 TTGAAGCAAGAGCGACCTCTTCTGCAAAACAACACACAGCTCCAAGC 1301
212  sgInAlaTyrAspLeu.....AlaLysAlaL 221
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1302 CAGGTGATGATGCTGGAGACTGCGAGGCGCCAGCAGGAGCCAAAGCTGC 1351
221  euLeuLysArg.....ThrAlaGlnAlaIleGluProTyr 232
|| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1352 TTGCTGAGCGGGCCACTTCGAAGAAGAAAGCAGCAGCTGTCTAGCCTG 1401
233  IleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSe 249
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1402 ATCACTGACCTG.....CAGAGCTCCATCTC 1427
249  rAspLeuSerGlu.....HisV 255
| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1428 CAACCTCAGCCAGCCAGCAAGAGAGCTGGAGAGCGCTCCAGGCTCATG 1477
255  alPheAspIleLeuLeuGluLeuTyrAsnIleAspSerHisLeu..... 269
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1478 GGGCCGGTTGACTGCCCGTGGCTCTCTGACCTCTGAGCTCACCACA 1527
270  LeuLeuSerValLeuProGlnLeuGluPheLys.....LeuLysSe 283
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1528 CTCATATGCCACCATCCAGCAACAGATCAAGAACTGGCTGGCTGAAGCA 1577
283  rAsnAspAsnGluGluAtgLeuGlnValValLysLeuLeuAlaLysMetP 300
:|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1578 GCAGGCCAAAGAGACAGCGGCCAGCTAGCACAGACCCCTC..... 1617
300  heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
:|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1618 .....CAACAGCAAGAAACAGGCTCCACAGGC..... 1644

```

```

317  CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1645 .....CTCCGCCACCAGGT 1658
333  sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
:|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1659 GGAGCAGCTAAGCAGTAGCTGAAGCAGAGGAGCAGCAGTTG...AAGG 1705
350  spLeuThrGluTyrLeuLysValArgSerHisAspProGluAlaIle 366
:|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1706 AGGTAGCGGAG.....AAGCAGGAGGCAACT 1731
367  ArgHisAspValIleValSerIleValThr..... 376
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1732 AGGCAGGACCATGCCAGCAACTGGCCACTGTCGAGAGGAGCAGAGGC 1781
376 ..... 376
1782 CTCCTTAGGGAGCGGATGCGGCTCTCAAGCAGCTGGAGCAGCTGGAGA 1831
377 .....AlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeu 390
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1832 AGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGCT 1881
391  AsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGl 407
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1882 AATGAAGCCCGGACAGTGCCAGACCTCAGTGACACAGGCCAGCGGA 1931
407  uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr...AlaLeuGlnS 423
| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1932 GAAG.....GCAGAGCTGAGCGGAAGGTGGAGGAACCTCCAGG 1969
423  erAlaAla.....GlyLysAspAlaAlaLysGlnIle 433
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1970 CTTGTGTTGAGACAGCCCGCAGGAACAGCATGAGGCCCGCCAGGT 2019
434  AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2020 GCAGAGCTAGAGTTGCTGCGGTGTGAGCAGCAAAAAGCAACTGAGAA 2069
450  pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
:|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2070 AGAAAGGTGGCCAGGAGGAGGAGCAGCTCCAGGAGCAGCTCCAGGCC 2119
467  snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2120 TCAAGAGTGCTTGAAGGTCAACAAG.....GGCAGC 2151
484  LeuAsp.....LeuAsnAlaValLysAlaLeuAsnGluMetTrpLys 497
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2152 CTTGAAGAGGAGAGCGCAGGCTGCAGATGCCCTGGAAGAGCAGCAGCG 2201
497  sCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIle 514
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2202 TTGTATCTCTGAGCTGAGGAGGAGACCCGAGCCCTGTTGGAGCAGCAT 2251
514  ysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
||| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2252 AGCGGGAACGAAAGGAG..... 2268
531  ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
2268 ..... 2268
547  sLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuG 564
||| :|:|:|:|:|:|:| :|:|:|:|:|:|:| :|:|:|:|:|
2269 .....CTGGAAGAGAGAGGCGCTGGGCCAAGAGGGCTGG 2302
564  luValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
||| :|:|:|:|:|:|:| :|:|:|:|:|:|:| :|:|:|:|:|
2303 AGGCTCATTTACTGCAGCTTGGGAGGCCCATCAGCTGAG..... 2343
581  ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPh 597

```



```

XX      15-JUL-1994   (first entry)
XX      Sequence encoding inner nuclear matrix (INM) protein MT2.
XX      Malignant cell type marker; interior nuclear matrix; MT2; NuMA; ss.
XX      Homo sapiens.
XX      OS
XX      Key          Location/Qualifiers
XX      CDS           1..6306
XX      FT            /*tag= a
XX      PN            WO9400573-A.
XX      PD            06-JAN-1994.
XX      PF            21-JUN-1993;    93WO-US06160.
XX      PR            22-JUN-1992;    92US-0901701.
XX      PA            (MATR-) MATRITECH INC.
XX      PI            Lidgard GP, Toukatly G;
XX      PS            WPI; 1994-026210/03.
XX      PT            P-PSDB; AAR47143.
XX      Novel malignant cell type markers of the interior nuclear matrix
XX      - used for detecting abnormal cell types e.g. malignant breast,
XX      prostate, lung, etc., for determining deg. of cell death in
XX      tissue, etc.
XX      Claim 24; Page 56-66; 93pp; English.
XX      MT1 and MT2 proteins were isolated from malignant cells essentially
XX      following the method of Penman and Fey, described in US Pat. Nos. 4
XX      882,268 and 4,885,236. DNA sequences encoding these proteins were
XX      cloned by probing expression libraries using monoclonal antibodies
XX      raised against the isolated INM proteins. The genetic sequence
XX      encoding MT2 has been disclosed by Yang et al. 1992, J. Cell Biol.
XX      CC 116:1395-1408, and is referred to by them as 'NuMA'. The nucleotide
XX      sequence encoding MT1 has not been described previously. Both the
XX      CC MT1 and MT2 proteins are distributed throughout the nucleus (with
XX      the exception of the nucleolus) in non-mitotic cells, and localise
XX      CC to the spindle during mitosis. A method of detecting an abnormal
XX      cell type in a sample which uses a mRNA complementary to
XX      AAQ54841 is claimed.
XX      SQ      Sequence 6306 BP; 1623 A; 1749 C; 2010 G; 924 T; 0 other;

alignment_scores:
Quality: 179.50 Length: 1514
Ratio: 0.277 Gaps: 60
Percent Similarity: 42.801 Percent Identity: 18.428

alignment_block:
US-09-512-581-2 x AAQ54841 ..

Align seg 1/1 to: AAQ54841 from: 1 to: 6306

16 ProProGluValLysGluIleSerAspLysIleSerLysGluMetVa 32
||| ::::|||||: ||||| ::::|||||: ::::
823 CCCAAGGACGTGGAGAGCTGCCTGACAAG.....AATGAGAGCCTTAC 866
32 lArqGlLeuLysMetValValLysThrPheMetAspMet..... 45
: |||||: ::::::|: ::::|:
867 CATGCCGTCGATCAAAACCCTGAAGACAGTGCCAGGACCTGAACACAGAGA 916
46 .....AspGlnAspSerGluGluLysGlu 954
||| ::::|||||: ::::|: |||||: ::::
917 AGAGCGAGATGATCGAAAATAACCATGCTTTTCGAGAGGAGTAATGAGAC 966

```

```

alignment_scores:
  Quality: 179.50      Length: 1514
  Ratio: 0.277         Gaps: 60
  Percent Similarity: 42.801  Percent Identity: 18.428

alignment_block:
  US-09-512-581-2 x AAQ54841 ..

Align seg 1/1 to: AAQ54841 from: 1 to: 6306

16 ProProGlyVallysGluIleSerAspLysIleSerLysGluMetVa 32
||| : : : : : : : : : : : : : : : : : : : : : : : : :
823 CCAAGGAGGTCCTGAGGAGCTGCCTGCAAG.....AATGAGAGCTTAC 866

```



```

398 uAspLysArgTrrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnI 415
3553 .....TCCGCATTAGCCACG 3568
415 leTyrlLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaLys 431
416 .....TCCGCATTAGCCACG 3568
3569 TT.....AGTATTGTTTGCACAAAGGACTTGGG.....CTTAAG 3603
432 GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSe 448
433 .....TCCGCATTAGCCACG 3568
3604 GAAGTCGATACAGCCCTTAGATGCTTGGCATCTGTTAGTTCGCCATCT 3653
448 rIleAspAspArgLeu.....LeuValGluArgIlePheAla..... 460
449 .....TCCGCATTAGCCACG 3568
3654 AAATGATGAAGAACTCTCTACTGTTATAGATAGCTTAATGTCATCATAC 3703
461 .....GlnTyrMetValProHis 466
3704 TTCAAAAGTGGTCTGAGTTCACCGAAAACTTCGAAATATAGTGTACAGT 3753
467 AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu..... 480
468 .....TCCGCATTAGCCACG 3568
3754 ATACTGGATACCTTAATCAAGAGAAATCGGACCTGATTTGAAATTAA 3803
481 ....TyrAlaThrLeuAspLeuAlaValLysAlaLeuAsnGluMet 496
482 .....TCCGCATTAGCCACG 3568
3804 ACCTTACACTACTTTGGCTTTAGTAGCAAGCCTGAATTAGGTATTTAG 3853
496 rpLysCysGlnAsnLeuLeuArg.....HisGlnValLysAspLeuLeu 510
497 .....TCCGCATTAGCCACG 3568
3854 CTCGGTATGGCCCAATTCGAAAGTGGTGAATAAATAAGAGTACCAG 3903
511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSe 527
512 .....TCCGCATTAGCCACG 3568
3904 GACCTTATA.....CCCATATTGCTATAACTTGAAAGT.....AG 3941
527 rLysValMetValIleThrArgAsnLeuProAsp..... 538
528 .....TCCGCATTAGCCACG 3568
3942 TAACAAGTATGTCATAAACCACAAATTTAGACGATATAGAGTATATCTTC 3991
539 .....ProGly 540
3992 GGAGAAGCAGACAGAAAGATCGATTGATTTTACACCAAGAGTTGGG 4041
541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspG 557
542 .....TCCGCATTAGCCACG 3568
4042 CAAACTTCTGATATAACATTAGTCTTTGGGTGCTTTATTAGACACATCTCA 4091
557 uLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysL 574
575 .....TCCGCATTAGCCACG 3568
4092 TAAGTTTAGAAATTTAGACAGGACCTATGCCGAGAAGTGCGCCAATGTA 4141
574 ysGlnAlaGluGlyCysValArgGluIleThrLysLys..... 586
575 .....TCCGCATTAGCCACG 3568
4142 TCAGTATGATGGT...GTTTAGAGGTTTACAAAGCATGAGTTTAAAGA 4188
587 .....LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMe 600
588 .....TCCGCATTAGCCACG 3568
4189 ACAACATATTCAGAAACGAAAGTTTATGATTTGAAATGATGATGTTCAAC 4238
600 tileLysPheLeuLeu.....GluArgIleAlaProValHisI 613
601 .....TCCGCATTAGCCACG 3568
4239 TATTAAAGTCTTGATATGGGTGCATAAATGATATCTCTGCTCGGTTTT 4288
613 leAspThrGluSerIleSer.....AlaLeuIleLysGln 624
614 .....TCCGCATTAGCCACG 3568
4289 GCCAAGTGAAGAAATCCCAAGCAAGTATGTTGTTGGCTTGTGCATACAG 4338
625 ValAsnLysSerIleAspGlyThrAlaAspAsp.....G 636
626 .....TCCGCATTAGCCACG 3568
4339 GAATCATTAATAATATTCGGGGCTAAGTTTCAGAGTATGGGATGAACCA 4388
636 .....TCCGCATTAGCCACG 3568
636 uAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuL 653
637 .....TCCGCATTAGCCACG 3568
4389 TAAAGAATTTATATCAAAATGAAGCCAAACTA.....TGGGAAAAGT 4429
653 euLysValLeuSerPheThr.....HisProIle.....SerPheHis 665
654 .....TCCGCATTAGCCACG 3568
4430 TTAACTCTGCTCTCCAGACCAACCACTATCCCTTTTATCTCTCTGTAT 4479
666 SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspG 682
667 .....TCCGCATTAGCCACG 3568
4480 CTTCGCAATCATCGAAAGAATATGTCGCG...CTAAATATCTCTCTAA 4526
682 uLysValAlaGluAla.....AlaLeuGlnI 691
683 .....TCCGCATTAGCCACG 3568
4527 TAACTTCAAGGAAGGATACCAAAATTTGGGTGAAAGGTTTACATTTGAT 4576
691 lePheLysAsnThrGly..... 696
692 .....TCCGCATTAGCCACG 3568
4577 TATTTGAAA...ACAGTACACAGAAATCATCCAGGACACGTTTTC 4623
697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProVa 713
698 .....TCCGCATTAGCCACG 3568
4624 TCTTTGATTAGGAAGATGATGGCTCACTATCAAAATTTTGTCTACTTA 4673
713 IleuHis.....HisLysSerLysLysGlyPro..... 723
714 .....TCCGCATTAGCCACG 3568
4674 TATTTCTCTGGACATTTATCAAGGCAGAAAAAGGAACTCCATACGCTG 4723
723 ..... 723
4724 ATATTTTAAACGGGATTTATTTGAATTTGACAGCATTTTTCAGTGCAT 4773
724 .....ArgGlnAlaLysTyrAlaIleHisCysIleHi 734
4774 CTGGAAGGAATGAATAACTTGCAGTGGATTCGTTAAGAATGCTGTATGA 4823
734 sAlaIlePheSerSerLysGlu.....ThrGlnPheA 745
735 .....TCCGCATTAGCCACG 3568
4824 ATCCATCTCTCAGAGTTCGAAATATTTGCAAAAATGGGCAACTGAGTTTA 4873
745 laGlnIlePheGluProLeuHis.....LysSerLeuAsp 756
746 .....TCCGCATTAGCCACG 3568
4874 ACAAAATTTACAGTAACATACACGGCACTTTTATCATTTAAAGATACGAAA 4923
757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773
778 .....TCCGCATTAGCCACG 3568
4924 ACAACTAACATG..... 4935
773 eAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValA 790
774 .....TCCGCATTAGCCACG 3568
4936 ....CTTTTGAGAATAGATGAGTTTTTGGGAAACACCCCTTCA..... 4974
790 laThrPheIleValLysAspLeuLeuMetAsnAspArgLeuPro..... 804
791 .....TCCGCATTAGCCACG 3568
4975 .....GATTTGCTAGCTCAACGCTCTCTAGACGCGAT 5007
804 ..... 804
5008 TCTTTTGAAGGCTGCTCTATATACCTTGAACAGTGCCTATCGACAGATCC 5057
805 .....GlyLysLysThrThrLysLeuTrpValProA 815
806 .....TCCGCATTAGCCACG 3568
5058 TCACGATAAGAACCAAAATGGCAACACTACTGAAAAATTTTACAAATCACA 5107
815 spGluGluVal..... 818
816 .....TCCGCATTAGCCACG 3568
5108 ACCAAGAAATAGGACACATTTGACTCACTCGATGGTGTACTGAGAACCTTT 5157
819 ...SerProGluThrMetValLysIleGlnAlaIleLysMetMetValAr 834
820 .....TCCGCATTAGCCACG 3568
5158 GCTACAGGAAGAACTTGGTTCTTAAATTTGAAGAATTTGCAATATTTCTCAAAA 5207
834 gTrp.....LeuLeuGlyMetLysAsnAsnH 843

```







```

4874 AACAAATTACAGTAACACTACACGCGACTTTTATCATTAAGATACGAAA 4923
757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773
4924 ACAACTAACATG..... 4935
773 eaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpVal 790
4936 ....CTTTTGAGAAATAGATGAGTTTGGGAAACAACCCCTTCA..... 4974
790 laThrPheIleValLysAspLeuMetAsnAspArgLeuPro..... 804
4975 .....GATTTGCTAGCTCAACGCTCCTTAGACGGAT 5007
804 ..... 804
5008 TCTTTTGAAGGTCCTCTATACCTTGAACAGTCTATCGACAGAAATCC 5057
805 .....GlyLysLysThrThrLysLeuTrpValProA 815
5058 TCACGATAAGAACCAAAATGGCAACTACTGAAAAATTTACAAATCACAT 5107
815 spGluGluVal..... 818
5108 ACGAAGAAATAGGAGACATGTGACTCAGTGGTGTACTGAGAACCTTT 5157
819 ...SerProGluThrMetValLysIleGlnAlaIleLysMetMetValAr 834
5158 GCTACAGGAAATCGTTTCTTAAATGGAAGTTCGCAATATCTCGAAAA 5207
834 gTrp.....LeuLeuGlyMetLysAsnAsnH 843
5208 CTGGAACCTCGCACAGACTGCTTTAATCTCTCGCAAAATTTTCAGATG 5257
843 isSerLysSerGlyThrSerThrLeuArg.....Leu 853
5258 ACCCCAAACTACACAGGATGCTTAAATCTCTCGCAAAATTTTCAGATG 5307
854 LeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysI 870
5308 TATTCTCAATATATTCGAACCTCTCGTTCATTTTCAGCGGAAAAAT 5357
870 eSer...LysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaI 886
5358 TTTCTTTGCTCCAGATGTGAAG..... 5379
886 leValLysLeuAlaGlnGluProCysTyrHisGluIleThrLeuGlu 902
5380 .....GAA 5382
903 GlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu...CysTyrGlnVa 918
5383 TGGTACAGCATAGGCTTCAAGCTGCAATCTAGAAGCAATGTTCAAC 5432
918 lArgGlnValPheAlaGlnLysLeuHis.....LysGlyLeuSerArgL 933
5433 TTTGAAATTTGGTAGAACAAATAGAGAGTTTAAAGAAATATTAGCATA 5482
933 euArgLeuProLeuGluTyr.....MetAlaIleCysAlaLeuCys 946
5483 GAGAAGTACTTTTGAGTACAAATATTGCGAAAGCTTAAATGCCATCTCA 5532
947 AlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVa 963
5533 AACGAGATCCATTAGGACTCAA..... 5556
963 lLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValS 980
5557 .AAATACATCCAAATTCCTTTAGGCTTATCGGAACAAATTTTAAACGT 5605
980 erGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996

```

```

5606 CATCTAAAGAGAGACGCGCTGCTAAAGAAAACAGAAATTTATGATGAAATTA 5655
997 HisLeuLeuAlaHisAspProAspTyrVal...LysValGlnAspIleG 1012
5656 CACAGTTTA...TATGACCTCAGTTTTTTATCTTCTCGGAAGATAGTT 5702
1012 uGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeu 1029
5703 TGAATACAAAGTAACACTACCATCTCGATTATCGAATGGAACGATTG 5752
1029 etAlaLysAsnGluAsnAsnSerHisAlaPhe...IleArgLysMetVal 1044
5753 GGGCTGACTTCGTGCAATCATTCATATTGTCAATGAGAAAGTCATTT 5802
1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMe 1061
5803 GACCAATTGAAATGAATGAACAAGCA.....GACGCTGACTT 5840
1061 tAsnGluLysLeuTyrThrValCysAspValAlaMetAsn..... 1074
5841 AGGAAAAACATTTCTCATTAGCCCAATTGGCGAGAAACAACGCTAGGC 5890
1075 ...IleIleMetSerLysSerThrThrTyrSerLeuGlu 1086
5891 TAGATATAGCCTCCGAATCATTAATGCAATGTTTGGAA 5928
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AA01274
seq_documentation_block:
ID AAX01274 standard; cDNA to mRNA; 8351 BP.
XX
AC AAX01274;
XX
DT 08-APR-1999 (first entry)
XX
DE Yeast ME1 coding sequence.
XX
KW Cell cycle checkpoint gene; yeast; radiation resistance; detection;
KW G2/M checkpoint; ss.
XX
OS Saccharomyces cerevisiae.
XX
FH Key Location/Qualifiers
FT CDS 784..7890
FT /tag= a
XX
PN US5866338-A.
XX
PD 02-FEB-1999.
XX
PF 06-JUN-1997; 97US-0870693.
XX
PR 18-FEB-1994; 94US-0198446.
PR 12-MAY-1992; 92US-0882051.
PR 14-MAY-1992; 92US-0884426.
PR 12-MAY-1993; 93WO-US04458.
PR 06-JUN-1997; 97US-0870693.
XX
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA (UNIV-) UNIV ARIZONA.
PA (UNIW) UNIV WASHINGTON.
XX
PI Groudine MT, Hartwell LH, Pion SE, Weinert TA;
XX
WPI: 1999-141919/12.
XX
P-PSDB; AAW73893.
XX
Nucleotide sequences that hybridise to huRAD cDNA - method for
isolating human checkpoint cDNA, and antibody to CDC34
XX
Example 4; Column 61-78; 73pp; English.
XX
This sequence encodes the yeast ME1 protein. The DNA sequence is
CC

```



```

539      ...ProGly 540
3992  GGAGAAAGCAGACAGAAAGATCGATTGATTTTACACCAAGAGGTGG 4041
541  LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspG1 557
4042  CAACCTCTGATAAACATAGTTTGGTGCTTTATTAGACACTTCTCA 4091
557  uLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysL 574
4092  TAAGTTTAGAATTTAGACAGGACCTATCGGAGAGGTGGCCCAATGTA 4141
574  ysGlnAlaGluGlyCysValargGluIleThrLysLys..... 586
4142  TCAGTATGATTGGT...GTTTAGACCTTACAAAGCATGAGTTTAAAAA 4188
587  .....LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMe 600
4189  ACAACATATTCAGAAAACGAAGTTTATGATTGAATGATAGTGTCAAAC 4238
600  tIleLysPheLeuLeu.....GluArgIleAlaProValHisI 613
4239  TATTAACTCTTGATATCGGTCATAAATGATATCCTCGTCTCGGTTTT 4288
613  leAspThrGluSerIleAspGlyThrAlaAspAsp.....G1 636
4289  GGCAAGGTGAGATCCAGCAAGCAATGTTGTTGGTGCCTTGTGCATAC 4338
625  ValAsnLysSerIleAspGlyThrAlaAspAsp.....G1 636
4339  GAATCATATAAATATTCGGGCTAAGTTCAGAGTCATGGGATATGAACCA 4388
636  uAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeu 653
4389  TAAGAATTATATCCAATGAAGCCAACATA.....TGGGAAAGT 4429
653  euLysValLeuSerPheThr.....HisProIle.....SerPheHis 665
4430  TTAACCTGTCTCCAAGACAACCATCTATCCGCTTTTATCTCTCTGTAT 4479
666  SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspG1 682
4480  CTTGCGCAATCATCGGAAGAATATGTCCG...CTAAATATCTCTCTAA 4526
682  uLysValAlaGluAla.....AlaLeuGlnI 691
4527  TAACCTCAAGGAAGATACCAAAATTTGGGTGAAAGGTTTACATGTGAT 4576
691  lePheLysAsnThrGly..... 696
4577  TATTGAA...ACAGGTACAACAGAAAATCATCAGGACACGTGTTTCC 4623
697  SerLysIleGluGluAspPheProHisIleArgSerAlaLeuProVa 713
4624  TCTTTGATTAGGAAGATGATGGCTCACTATCAAAATTTTGTGACTTAA 4673
713  lleuHis.....HisLysSerLysLysGlyPro..... 723
4674  TATTCTCTGGACATATTATCAAGGCAGAAAAGGAACCTCAACGCTG 4723
723  ..... 723
4724  ATATTTTAAACGGGATTATTATTGAATTTGACAGCATTTTCACTGCAAT 4773
724  .....ArgGlnAlaLysTyrAlaIleHisCysIleHi 734
4774  CTGAAGGAATGAATAACTTCCAAGTGGATCGTTTGAAGAATGTGCTATGA 4823
734  salalIlePheSerLysGlu.....ThrIlePhe 745
4824  ATCCATCTCTACAGTTTCCTGAATATTCGAAAATATGGCAACCTGAGTTA 4873
745  laGlnIlePheGluProLeuHis.....LysSerLeuAsp 756

```







```

971-u.....TyrLeuLysGlnHisA 977
:          |||||
2123 AGAATGGGGTCTGGGAGAGAGAGTTGAGAGGATATTAAACAGCCAT 2172
977 laalavalserGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
:      |||||
2173 CCACAGAGTGTACCTATTCTATCTAGCCCTTGGCAACGTGTGGCTC 2222
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
|||:|||||
2223 CAAACTTTACATCAGCCCGGAGATCGAGAAAAGAAAGAGCGTCTCA 2272
1010 pilleGluGlnLeuLysAspValLysGluLysLeuTyrPheValLeuGluI 1027
:      |||
2273 AGATCGTGTCTGCGCATCTACAACAAGTACTCAGAAATGATGCAAGA 2322
1027 leuMetAlaLysAsnGluAsnAsnSer.....HisAlaPhe 1039
|||
2323 ATCTGTATGCTGCCAATGGCATAGGAGCTGTTTGGCCCAACAAAGGATAT 2372
1040 IleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPr 1056
|||:|||||
2373 TTTCGTAAGCTCGTGTATGTTGCCAAGTAAGAGAACAGACAGAGA 2422
1056 o...AspAlaLysMetAsn..... 1062
:      |||
2423 TATTAGTGTGTGCGCTGAACCTAGCACACATCTATGTGGACAAAGC 2472
1063 .....GluLysLeuTyrThrValCys..... 1069
:      |||
2473 AGTACATCAGCGCGTTCAGATGTGAAACCTGCCTCCGAAAGTCTAT 2522
1070 .....AspValAlaMetAsnIleIleMetSerLysSerTh 1081
:      |||
2523 AGCACAAAACACTGAAGTTGTACTATTGTGGCCGCGCCCTCTTCAA 2572
1081 rThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhep 1098
:      |||
2573 GTGTGGCAAGTTACAGGAATGCAACAGACTTTGCTCAAGGCTAGACATG 2622
1098 heThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
:      |||
2623 TGCACCCAGTATACAGTTCTTATGTTTAATGTGCCTTGGTCCGCGAA 2672
1115 GluMet...LysSerPheThrProGlyLysProLysThrThrAsnVa 1130
:      |||
2673 AGATTAGTACCTCTGCTGCGAAGATGAAAAAAGTAATCTGAAGGAAGT 2722
1130 IleuGlyAlaValAsnLys.....ProL 1138
|||:|||||
2723 ACTTAATGCTGTCAAGAAGTGGAGCTTGCACATAGATACTTCAGTTATT 2772
1138 euSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
|||:|||||
2773 TGAGTAAAGTGGAGATAAATGAGATTTGATTGGCCCTTGCTGTACA 2822
1155 ValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLy 1171
:      |||
2823 GAAGCCAGGCGAGTGTCTGACTTACTGAGCCAGGCCCGCAGTACCATTGTG.. 2870
1171 sGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrT 1188
:      |||
2871 .....GCCCGGCACGCAACAAAGATGAAGAGCGGGAG..... 2906
1188 hrMetSerSerProLeuProGlyLysSerAspLysArgAspSer 1204
:      |||
2907 .....CTGGGCGCAAGCAAGAGAGCAAGAAAG 2933
1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysThrPr 1221
:      |||
2934 GAGCTGTTAAGGCAGAACTCTTAAGAACAGGAGAAAGACGTCTCAG 2983
1221 oValThrGluGlnGluLysLeu..... 1229

```

```

|||:|||||
2984 AGAAAAGGAAGACAAAAGAAACTTTTGGACAGCGGCCCGCCAGTATGTGG 3033
1230 .....GlyMetAspAspLeuThrLys 1236
:      |||
3034 AGAAGACCAAAATATTTCTTATCTTTACTGGTGAGACTGAAGCAACAAA 3083
1237 LeuValGlnGlnLysProLysGlySerGlnArgSerArgLysArgG1 1253
:      |||
3084 ...GAGAAGAAAAGAGGTGGTGGTGGAGCGGCTTCTAAGAAGGGAGG 3130
1253 yHis..... 1254
:      |||
3131 AGATTGATGAATTTGTCATGATGACACTGATGATGACCTACCTATAT 3180
1255 .....ThrAlaSerGluSerAsp 1260
:      |||
3181 CAAAAGAAAGAGAAGAAAGGTAGTGGCAGTGAACAAGAGGTGAA 3230
1261 GluGlnGlnTrpProGluGluLysArgLeuLys..... 1271
:      |||
3231 GATGAGAGGGTGGTGAGAGAAAGAAAAGAGAGAGAGACATCCAAA 3280
1272 ...GluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysL 1287
:      |||
3281 GGGAGAAGAGGATCTCATGATGATGAACAGAAATGGCCCC.....A 3324
1287 ysglyLysArgGlyArgProProLysProLeuGlyGlyThrProLys 1303
|||:|||||
3325 AACCAAAAACGACGTCCACCAAAAGCAGAGAAAGAAAGGGTCCCAAG 3374
1303 ..... 1303
3375 CCAGAACGTCTGCTCCATCAATGAAGGGGAAAATAAAATCCAAAGCCAT 3424
1304 .....GluGluProThrMetLysThrSerL 1312
:      |||
3425 AATTTCATCAAGTATGACACTTTCGGATGAGGATAAACTTAAATTCCTG 3474
1312 ysllysGlySerLysLysSerGlyProProAlaProGluGluGluGlu 1328
|||:|||||
3475 ATGAAGACATCCCAAGCAACAGCAACAGTGTACTCAGACGAGGAC 3524
1329 GluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysG1 1345
|||:|||||
3525 GAACACAGAAAAGAAATGTGCTCATCAGAGAGTATTCGATGAGAACCA 3574
1345 nHisArgValSer.....ArgA 1351
:      |||
3575 GAACAAGTCTGGCAGCGGCGCGCAGTCCCGGAGGCCACGAAGACAGC 3624
1351 rgAlaGlnGlnArgAlaGluSerProGlu..... 1360
|||:|||||
3625 GGTCAATCAGGACTCAGACAGTACCAGCCATCCAGAAAAGAGAGGCC 3674
1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyAr 1377
|||:|||||
3675 TCCGGTCTGAGCAGTCTGACAATGAATCTGTGCAGTCAGGGAGAGCCA 3724
1377 gProSerLysThrProSerProSerGlnPro 1387
:      |||
3725 CTCAGGAGTTTCTGAGAACCACTCTGCCCA 3755

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT91039

seq\_documentation\_block:

ID AAT91039 standard; cDNA to mRNA; 8351 BP.

XX AAT91039;

XX 25-FEB-1998 (first entry)

XX Yeast checkpoint control gene MEC1 cDNA.



```

609 ...TATAGAGACCATGAAAGACAGACAGATGCTTCTGGATACATTGGC 655
459 eAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysC 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
656 AGCGTATTATCTACAACAGCGTCGGAAGAAAAGAAAGAAAGGACATAAA 705
476 ysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeu 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
706 AGGATCTTATTACACAGGCCACCTTG...TTGTATACAATGGCCGATAAA 752
493 AsnGluMetTrpLysCysGlnAsnLeuArgHisGlnValLysAspLe 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
753 ATTATTATGATGATCAGAACCTTGTGGAGAGCGCTGCTTCTGCCT 802
803 ACTTGAGGCTGACAAATGGATCAAGCTGATGCACAGTTTCATTGTTGAC 852
521 .....SerValLysAlaIlePheSerLysValMetVal 531
853 TCATTCAGCTCCAAATAATATTCAGCGCTTCTTGGTAAAGCT...TGC 899
532 IleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLy 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 ATTTCCTTCAACAAGAGATTACAGAGCGCTTGTCTTACTATAGAA 949
548 sPheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluV 565
950 A..... 950
555 aLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArg 581
951 .....GCATTGCGTACTTAACCCAGAGTGTCCAGCG 980
582 GluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLe 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
981 GAAGTCGTTAGGAATGGGT.....CATTCCTTGT 1012
598 uGluMetIleLysPheLeuGluArgIleAla.....ProValH 612
1013 GAACTTAACAACCTGGAAGAGCTCGTGGCATTCAGCAGAGCGCTGG 1062
612 IsIleAspThrGluSerIleSerAlaLeuIle.....Lys 623
1063 AACTCAATTCCAAATGCGTGGAGCATTTGGTGGACTGGCTGTTCTAGAA 1112
624 GlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVa 640
1113 CTCACATAAAGAGCGTGT..... 1133
640 lProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLys..... 654
1134 .....TCCATTAATAAATGGTGTCCAGCTTCTTCCAGAGCT 1170
655 ...ValLeuSerPheThrHisProIleSerPhe...HisSerAlaGluThr 669
1171 ATACTATTGATCTAGCAACCTTGGATTGTAACCATTTGGCAAATCAC 1220
670 PheGluSerLeuLeuAlaCysLeuLysMetAspGluLysValAlaGl 686
1221 TTT.....TTCTCAAAAGGATTATAGTAAGTCCAGCA 1255
686 uAlaAlaLeuGlnIlePheLysAsnThr..... 695
1256 TCTGGCCCTCCATGATTCCTCAATAATACAGAGTGAAGCTATGCAAGCAG 1305
696 .....GlySerLysIleGluGluAspPhePro 704
1306 AGAGCTGTATCAGCTAGATCATTCATTCATGTTTCAGGAAGATTATGAC 1355
705 HisIle.....ArgSerAl 709
1356 CAAGCTTTTCTAGTACTATATCAAGCCACACAGTTTGGCTCATCTCTTT 1405

```

```

709 aLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnA 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1406 TGTGCTCCCATTT..... 1418
726 laLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysLysGluThr 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1419 .....TTTGGTTTGGACAAATGTATATTTATCGAGGTGACAAAGAAAT 1463
743 GlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAs 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1464 GCA...TCTCAGTCTTTGAGAGGTTTGAAGCTTAT...CCTAATAA 1507
759 nLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuL 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1508 TTACGAA.....ACTATGAAAATTCCTCGGCTCTCTCTATGCTG 1545
776 uAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
1546 CCTCAGAAGATCAAGAAAACAGAT..... 1571
793 lLeValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrTh 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1572 ATTGCCAAGGCCCATTTG.....AAGAAGTTCAC 1600
809 rLysLeuTrpValProAspGluGluValSerProGluThrMetValLysI 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1601 AGACAGTAT...CCCGATGAT.....GTTGAAGCTTGGATTGAAT 1638
826 leGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsn 842
1639 TGGCACAATCTTAGAACGACT.....GATATACAGGGTGCC 1676
843 HisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHi 859
1677 CTTTCAGCTATGGAAACAGCAACA...CGAATCCTTCAGGAGAAAGTGA 1723
859 sSerAsp..... 861
1724 GGCCGATGTTCTCCAGAGATTCTCAATAATGGGTGCCCTCCATTTTA 1773
862 .....GlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
1774 GACTTGAACCTTAGGGAGGCTAAGAATATTTTGGCGTCATGGAC 1823
877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluPr 893
1824 CGTGCAAAAGCAGAGCGGAACAT.....GATGAGCA 1855
893 oCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuA 910
1856 TTACTATAAACCCATTTCCTGTTACCACGTCATATATAATCTCGCCAGGCTA 1904
910 laIleAsnAspGluCysTyrGlnValArgGlnValPhe.....AlaGln 924
1905 .....TATGAGCGATGTGTGAATTCATTCATGAAGCAGAA 1937
925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlu..... 938
1938 AAACGTATATAAAACATCTTACGC.....GAACATCCCTAA 1972
939 .....TyrMetAlaIleCysAlaLeuCys..... 946
1973 TTATGTTGACTGCTATTTCCGCTAGGAGCCATGGCTAGAGATAAGGAA 2022
947 .....AlaLysAspProValLysGlu.....ArgArg 955
2023 ACTTTTATGAGCTTCAGATGGTTTAAAGGAAGCTCTTCAGATTATATCAG 2072
956 AlaHisAla...ArgGlnCysLeuValLysAsnIleAsnValArgGln 971
2073 GATCATCCAGATGCTTGTCTTTGATTGGCAATCTTCATTTGGCAAAACA 2122

```



1199	.....MetSerSerProLeuProGlyLysLys.....	1197
738	GGCCCGCCGCCAGCGGAGGCTCTCTGGGAGGACGCCAAGAAGAAGG	777
1198	.....SerAspLysArgAspSerAspLeuValArg	1208
778	TGCCATCAGCCTCGGATTACAGCTCAAACCCGACTCAGAT	818
1209	SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluGlu	1225
819	.....GGGGCCAAAGGAGGAGCCTGTGTTTACAGC	847
1225	ngluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGln	1242
848	ACAG.....	851
1242	ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu	1258
852	.....CGGTCCTCCATCTCCTCGTGGTCGCTCATCATCATCTCCTCAGAC	899
1259	SerAspGluGlnGlnThrProGluGluLysArgLeuLysGluAspIleLe	1275
900	TCAGATGTGTCT.....	911
1275	uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA	1292
912	.....GTCAAGAAACCCCTCGAGGTA	933
1292	fg.....ProProLysProLeuGlyGlyThrPro	1302
934	GAAGCCAGCTGAGAACGCCACCCCAAAACCC...CGAGGCGGAGACCA	980
1303	LysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys	1319
981	AAGCAGAACAGCACCACTTCACCTCCAGCAGTACAGTGCAGTGCAGC	1030
1319	rgLysPro.....	1321
1031	TGGTGAAGTAGACCGCATCACTGAATGGAAGAGACGTGATGAAGACGCA	1080
1322	.....ProAlaProGluGluGluGluGluArgGlnSer	1334
1081	GGCGTGACTCGGAGCTCGGAGCGCGGAGCAGGAGGAGGAGGAGCTCCGG	1130
1335	GlyAsnThrGluGlnLysSerLysLysGlnHisArgValSerArgAr	1351
1131	AGACTTCAGAAACAGGAGGAGGAGGAGGAGCGCGC.....	1169
1351	gAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrG	1368
1170	.....AAGGAGCGGCAGACGCTGGGGCAGCAGTGGAGAGGAGCTGGAGG	1215
1368	InSerThrPro.....GlnLysGlyArgGlyArgProSer	1379
1216	ACGAGGACCTGTGAAGACGCTAGCCGAAGGCTCGAGGCCGAGGCACA	1265
1380	LysThrProSerProSerGlnProLys	1388
1266	CCATCTCTCTGACTCAGACCTTGA	1292

seq\_name: /SIDS1/qcdata/geneseq/geneseq/NA1995.DAT:AA086236

seq\_documentation\_block:

seq\_documentation\_block:  
ID: AAQ86236 standard; cDNA to mRNA; 4548 bp.

AAQ86236  
AAO86236:

DT 30-NOV-1995 (first entry)

Megakaryocyte stem cell maturation factor clone 1204B.

XX	OS	Homo sapiens.	
XX	OS		
XX	Key	Location/Qualifiers	
FT	CDS	633..3908	
FT		/*tag= a	
XX	XX		
PN	JP07067658-A.		
XX	XX		
PD	14-MAR-1995.		
XX	XX		
PF	03-SEP-1993;	93JP-0243767.	
XX	XX		
PR	03-SEP-1993;	93JP-0243767.	
XX	XX		
PA	(MOCH ) MOCHIDA PHARM CO LTD.		
XX	XX		
DR	WPI; 1995-143852/19.		
XX	P-PSDB; AAR70982.		
DR			
XX			
PT	Novel DNA involved in differentiation of stem cells to		
PT	mega:karyocyte(s) - used in diagnosis and treatment of diseases		
PT	involving abnormal differentiation, e.g. megakaryocytic leukaemia		
XX			
PS	Claim 1; Pages 18-23; 41pp; Japanese.		
XX			
XX			
CC	AA086236 encodes AAR70982 the megakaryocyte (MKC) stem cell maturation		
CC	factor clone 1204B. The DNA is involved in the differentiation of		
CC	stem cells to MKCs and platelets, it can be used as a probe to		
CC	diagnose diseases involving abnormal differentiation, it may also		
CC	be used in the treatment of MKC leukaemia. Antibodies raised		
CC	against the protein can be used to immunologically measure the MKC's		
CC	level of differentiation.		
XX			
SO	Sequence 4348 BP; 1431 A; 923 C; 1096 G; 1098 T; 0 other:		

seq\_documentation\_block:

seq\_documentation\_block:  
ID: AAQ86236 standard; cDNA to mRNA; 4548 bp.

AAQ86236  
AAO86236:

DT 30-NOV-1995 (first entry)

Megakaryocyte stem cell maturation factor clone 1204B.



1385 rGlnProLysLysAsnVal 1391  
 :::::::::::::::  
 2189 AGAACCAAGCGCAGGTG 2207

seq\_name: /SIDS1/gcgdata/geneseq/NA1998.DAT:AAV01731

seq\_documentation\_block:

ID AAV01731 standard; cDNA; 2108 BP.

AC AAV01731;  
 DT 27-MAR-1998 (first entry)  
 DE Mouse liver cancer-originated culture cell growth factor encoding cDNA.  
 KW mouse; human; liver cancer-originated culture cell growth factor;  
 KW HHDGF; HET-A; HET-B; ds.  
 XX Mus sp.  
 OS Mus sp.  
 FH Key Location/Qualifiers  
 FT CDS 99..2108  
 FT /\*tag= a  
 FT /product= "Liver cancer-originated culture cell growth  
 FT factor"

XX JP09313185-A.

PN 09-DEC-1997.

XX 27-MAY-1996; 96JP-0131788.

XX 27-MAY-1996; 96JP-0131788.

XX (KISHI/) KISHIMOTO C.

PA (SEKI ) SEKISUI CHEM IND CO LTD.

XX WPI; 1998-080076/08.

DR P-PSDB; AAW37483.

XX DNA segment encoding protein homologous to human liver

PT cancer-originated culture cell - which may be modified to produce

PT polypeptide of at least 5 continuous amino acids, useful in

PT producing, e.g. protein HET-A

XX Claim 9; Page 10-12; 18pp; Japanese.

XX The present sequence encodes a protein which has local homology to a  
 CC human liver cancer-originated culture cell isolated from a mouse  
 CC testicle cDNA library. The sequence may (1) be modified to produce a  
 CC polypeptide comprising at least 5 continuous amino acids; (2) a  
 CC polypeptide protein similar to protein of (1), but comprises no protein  
 CC combined naturally and also has 2 amino acid sequences (as given in  
 CC the specification) which may be modified to produce a sequence of at  
 CC least 5 amino acids; (3) a recombinant DNA molecule comprising a vector  
 CC and the DNA segment of (1); (4) a cell comprising (3); (5) producing a  
 CC polypeptide comprising amino acids corresponding to a protein of (1),  
 CC in which a cell comprising (3) is cultured under a conditions enabling  
 CC the expression of (1) to produce the polypeptide which is isolated.  
 CC The above method may be used to produce proteins HET-A and HET-B.

XX Sequence 2108 BP; 567 A; 557 C; 681 G; 302 T; 1 other;

alignment\_scores:

Quality: 180.00 Length: 543  
 Ratio: 0.800 Gaps: 23  
 Percent Similarity: 41.436 Percent Identity: 20.994

alignment\_block:

US-09-512-581-2 x AAV01731 ..

Align seg 1/1 to: AAV01731 from: 1 to: 2108

931 SerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAl 947  
 ||| :::::::::::::::  
 20 TCAGCGTCCGACTCCGCTCTGACTTCCGCTCTCACCGCGTGTGTCG 69  
 947 alyaspProValLysGluArgAlaHisAlaArgGlnCysLeuValL 964  
 ||::: :::::::::::::::  
 70 TTCCCGCGCGCTCGGCTCCGCGCAGCATCGCGCAGCCTTCAAGCCC 119  
 964 ySAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980  
 :::::::::::::::  
 120 GGGGACTTGGTGTTCCTCAAGAT. .... 142  
 981 GluLysLeuLeuSerLeu. .... 986  
 ||| :::::::::::::::  
 143 GAAGGCTACCCGCACTGCCGCGCGGATTGATGACATTGCTGATGGTG 192  
 987 .....LeuProGluTyrValProTyrThrIleHisLeuLeu.Ala 1000  
 ||| :::::::::::::::  
 193 CCGTGAAGCCGCCCAACAAATACCCCATCTTCTT...CTTGAACG 239  
 1001 HisAspProAspTyrValLysValGlnAspIle. .... 1011  
 ||::: :::::::::::::::  
 240 CATGAACCGCCTTCTCGGACCCCAAGGACCTTCCCTATCATGATAGTG 289  
 1012 .....GluGlnLeuLysAspValLysGluCysL 1021  
 ::::: ||| :::::::::::::::  
 290 CAAGGACAAGTACGGAAGCCCAACAAGAGAAAGCTTCAATGAGGGC 339  
 1021 euTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHis 1037  
 |||| :::::::::::::::  
 340 TCTGGGAGATC. ....CAGAACAACCCCAT 365  
 1038 AlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaG 1054  
 ||| :::::::::::::::  
 366 GCC. ....AGCTACAGTCGCGCTCCGCGGTGAGCTCTCTGCACATGA 409  
 1054 nGlyPro. ....AspAspAlaLysM 1061  
 ::::: ||| :::::::::::::::  
 410 GCCCCTGAAGCGCAGCTGGTGTGGCAGCGACGTGGACAGGACAAAG 459  
 1061 etAsnCluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMet 1077  
 :::::::::::::::  
 460 AATCCCGAGGGTCATGACTGTG. ....ACAGCTGTGACC 494  
 1078 SerLysSerThrTyrSerLeuGluSerProLysAspProValLeuPr 1094  
 ::::: ||::: :::::::::::::::  
 495 ACCACAGCCACCACTACACAGGATGGAGCGGATTCCTGAC. .... 533  
 1094 oAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnT 1111  
 ||| :::::::::::::::  
 534 .....TCTCACAAGAGCAGTATCAGT. .... 557  
 1111 yrLeuProGluMetLysSerPhePheThrProGlyLysProLysThr 1127  
 ::::: ||| :::::::::::::::  
 558 .....GGCCTGAAGCGGAAG 572  
 1128 ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysG 1144  
 ||| :::::::::::::::  
 573 ACACCACTTGAAGTG. ....TCAGTCTCTAAACG 604  
 1144 nSerGlnThrLysSerArgMetGluThrValSerAsnAlaSerSerS 1161  
 ::::: ||||| :::::::::::::::  
 605 AGCTAGACAGAGCTTCCAGTCAGCTGGATCAGGCCAGTGTGTCCTCCATCAG 654  
 1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSer 1177  
 :::::::::::::::  
 655 AAGAGGAGCTCTGAGAGCCCA. ....TCT 677  
 1178 GluMetAspHisSerGluAsnGluAspTyrThr. .... 1188  
 ||| :::::::::::::::  
 678 GAGTCGGAGAAGACCAGTCACCAGGATTTTACCCCAAGAGACAGCAGC 727







1739	GAAGCTTTTACCAGCACTGGCTTAATAAAAGCTTCTCGAAGAACACAGA	1788
1287	ysGlyLysArgGlyArGProProLysProLeuGlyGlyThr.....	1301
1789	AGGGAAAAGAGGAGGAGGAGCAGGAATCAGGGGAGCACACTCAGGTT	1838
1302	ProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLy	1318
1839	CCAGCGATTCTCCGCACACGCCAGGAGGAGCAAAGGGCGAGAGCTCGC	1888
1318	sSerGlyProProAlaProGluGlu.....	1326
1889	CTCATCCCTCGAGGACCCGAGAGATCAGTGCTCTCGAAAGGGCTTAG	1938
1327	.....GluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys	1340
1939	CCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTC CGATGGA	1988
1341	SerLysSerLysGln.....	1345
1989	GAGAAAAAAGAGAGAGGTGTCACTCCCTGGGCATCATTC AAAAAGATGGT	2038
1346	.....HisArgValSerArgArgAlaGln.....	1353
2039	GACGCCAAAGAGCGTGTTAGACGGCTTCGGAAGTGATAAAGAGATG	2088
1354	.....GlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln	1368
2089	AGCTGGACAAGGTCAAGAGCGTACTCTGTCTCCACCGAGAGCACAGCC	2138
1369	SerThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSe	1385
2139	TCTGAATGCAAGACAAATGAAAGGAGCGTGGAAAGAGCCAAAGCCGA	2188
1385	rGlnProLysLysAsnVal	1391
2189	AGAACCAAGCGCAAGGTG	2207
 seq_name= /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA74903		
seq_documentation_block:		
ID	AAA74903 standard; cDNA; 6605 BP.	
XX		
AC	AAA74903;	
DT	26-JAN-2001 (first entry)	
XX		
DE	Human gravin coding sequence.	
XX		
KW	Human; gravin; PKA RII binding site; myasthenia gravis;	
KW	kinase anchoring protein; cAMP dependent protein kinase; ss	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	192..5534
ET	/tag=a	
ET	/product="gravin"	
XX		
PN	USG090929-A.	
XX		
PD	18-JUL-2000.	
XX		
PF	19-DEC-1997; 97US-0994570.	
XX		
PR	19-DEC-1996; 96US-0769309.	
XX	(UYOR-) UNIV OREGON HEALTH SCI.	
PA		
XX		
PI	Klauck TM, Scott JD, Nauert JB;	
XX		
DR	WPI; 2000-523763/47.	



```
5291 ATCATTCAGGCTGACACACTTTTGGATGAATCAGAGAAAAGAACCCCA 5340
1114 OGU.....MetLysSerPheThrProGlyL 1124
5341 GCAAAAGAGAGCGTCTTAAGCGTTTAAAGGAGCAACTGAATGACATAC 5390
1124 ysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
5391 GCCAAGGTGGAGCTCTACACGTGACCAAGCAACAACCTGATGGCAAC 5440
1141 Alagly.....LysGlnSerGlnThrLys 1149
5441 CGCGGTGACCACTGCAGGAAATTTAGTAGCCCAAAATCTCAGAGCTCAA 5490
1149 rSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerS 1166
5491 CCATCGATTTCAGCCATTTTCACACAGAAATTAAGACTGGAAAGCCCTCA 5540
1166 erProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSer 1182
5541 TTCT.....TTGAAGGAATTTGGAGCAGTTT 5566
1183 GluAsnGluAspTyrThrMetSerSerProLeuProGlyLys..... 1196
5567 AACTCAGATATACAAAATTTGCTTGAACCACTGGAGCTGAAATTCAGCA 5616
1197 .....LysSerAspLysArgAspAspS 1204
5617 GGGGGTGAATCTGAAAGAGGAGCACTTCAATAAGATATGAATGAAGACA 5666
1204 erAspLeuValArgSerGluLeuGlyLysProArgGlyArgLysLysThr 1220
5667 ATGAGGCTACTGTAAGAAATTTGTCAAAGAGGAGGACAACTTACAACA 5716
1221 ProValThrGluGlnGluLysLeuGlyMetAspLeuThrLysLe 1237
5717 AGAATCACAGATGAGAGAGAGAGAGGAATAAAGATAAAGACAGCAGCT 5766
1237 uValGlnGluGln.....LysProLysGlySerGlnArgSerA 1250
5767 GTTACAGACAAAACATAATCTCTCAAGGATTTGAGCTCTCAAGAAGAA 5816
1250 rLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGlu 1266
5817 AAAG.....GCTAGAAATTTCTCATCAGTGTGATCAG 5851
1267 GluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSe 1283
5852 TACAAGAGGCGAGCTGATGATCTCTGAAATGCTTGGATGACATT..... 5896
1283 rProProLysGlyLysArgGlyArgProProLysProLeuGlyGlyG 1300
5896 ..... 5896
1300 LysThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
5897 .....GAAAA 5902
1317 LysLysSerGlyProAlaProGluGluGlu.....GluGluGlu 1330
5903 AAATTAGCCAGCTACCTGAGCCAGGAGGAAAGGAAATAAGGAAT 5952
1330 uGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisA 1347
5953 TGATCGGGAATG.....CAGAAGAAGAAAGAGGAGCTGAATG 5990
1347 rgValSerArgAlaGlnGlnAlaGluSerProGluSerSerAla 1363
5991 CAGTGCGTAGCAGCTGAGCGGCTGTCTGAGGATGGCGCCGCAATGGCA 6040
1364 IleGluSerThrGln 1368
:::|||||
```

6041 GTGGAGCCAACTCAG 6055

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV23545

seq\_documentation\_block:

ID AAV23545 standard; cDNA; 6605 BP.

XX AAV23545;

XX 13-JUL-1998 (first entry)

XX Human gravin polypeptide coding sequence.

XX Gravin: kinase anchoring protein; type II regulatory subunit; PKA; PKC;  
XX CAMP-dependent protein kinase; protein kinase C; autoimmune disease;  
XX Myasthenia gravis; nicotinic acetylcholine receptor; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 192..5534  
XX /\*tag= a

XX US5741890-A.

XX 21-APR-1998.

XX 19-DEC-1996; 96US-0769309.

XX 19-DEC-1996; 96US-0769309.

XX (UYOR-) UNIV OREGON HEALTH SCI.

XX Klauck TM, Nauert JB, Scott JD;

XX WPI; 1998-260552/23.

XX P-PSDB; AAW53863.

XX New polypeptide fragments of protein kinase binding protein gravin -  
XX are useful for the study of modulation of action between gravin and  
XX protein kinase(s)

XX Example 1; Column 19-32; 32pp; English.

XX This sequence encodes the human gravin polypeptide, which represents  
XX a polypeptide of the invention. The polypeptides are fragments capable of  
XX binding to type II regulatory subunit of CAMP-dependent protein kinase  
XX (PKA). Gravin is a kinase anchoring protein that binds to type II  
XX regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an  
XX antigen of the autoimmune disease Myasthenia gravis (MG), where a patient  
XX develops antibodies against their own nicotinic acetylcholine receptors.  
XX The polypeptides are useful for providing analogues of gravin in the  
XX study of the modulation (e.g. blocking, inhibiting and stimulating) of  
XX interactions between gravin and kinase. The peptides are involved in the  
XX modulation of gravin-kinase interactions.

XX Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;

alignment\_scores:

Quality: 181.00 Length: 473  
Ratio: 0.834 Gaps: 19  
Percent Similarity: 45.877 Percent Identity: 20.930

alignment\_block:

US-09-512-581-2 x AAV23545 ..

Align seg 1/1 to: AAV23545 from: 1 to: 6605

1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052

|||||

930 AGCCACGCAAGAAATTTCTCCCGGAGCGAATCTGGCCCAAGCAGTGGAGGA 979







XX Muscular dystrophy gene - used for prepn. of probes, dystrophin  
 PT polypeptide and antibodies for diagnosis and therapy of muscular  
 PT dystrophy  
 XX  
 PS Disclosure; Fig 5; 69pp; English.  
 XX  
 CC The inventors claim an MD probe comprising a purified ss NA SQ which  
 CC hybridises to at least a part of the MD gene; pure dystrophin (DS)  
 CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The  
 CC probes are equal to or greater than 10b of one of 12 cDNA sequences  
 CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd  
 CC gene.  
 XX  
 SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T; 7 other;

## alignment\_scores:

Quality: 181.50 Length: 1405  
 Ratio: 0.305 Gaps: 57  
 Percent Similarity: 42.420 Percent Identity: 17.722

## alignment\_block:

US-09-512-581-2 x AAN90338 ..

Align seg 1/1 to: AAN90338 from: 1 to: 12923

```

110 IleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAs 126
111 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2804 ATAAAGTCAGTTAAATAATTTGTAAGAT.....GAAGTCAA 2841
126 nArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
127 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2842 CCGCTATCAGTCTCAACTCAATTTGACGATTAATAATTCAAAGCA 2891
143 leCysPheGlu.....LeuGluAspSerAsn 151
144 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2892 TAGCCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2941
152 GluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnG 168
153 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2942 GTGGCCTTTACAAATCATTTTAAAGCAAGTCTTTCTGATGTGAGGCCAG 2991
168 yHisAsnGln.....LysValHisMetHisMetV 178
169 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2992 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3041
178 alAspLeuMetSerSerIle..... 184
179 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3042 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3091
185 .....IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVa 199
186 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3092 CTCCTCCATACCTCAACTAGTGTGACGAGTATGAAATCATGGAGAGAG 3141
199 lLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrA 216
199 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3142 ACTCGGGAATTCAGGCTTTACAAAGTCTCTGCAAGAGAGAGAGAGAG 3191
216 spLeu.....AlaLysAlaLeuLeuLysArgThrAla 226
217 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3192 GCCTATATCTCAGCAGCAGTGTGAAGAGAGATGTGGAAGAGAGAGAG 3241
227 GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLe 243
228 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3242 TCTGAAATAGCCGGAATATCAATCAGAAATTTGAAGAAATTTGAGGAGC 3291
243 uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu... 258
244 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3292 CTGGAAGAGAGCTCTCTCCAGAGTGTGTTGAGCATTTGCAAAAGTAGAGG 3341
259 .....IleLeuGluLeuTyrAsnIleAspSerHisLeu.....LeuLeu 271
260 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

3342 AGCAAAATGAATAAATCCGAAAAATTCAGAAATCAGATACATAACAACTCGAAG 3391
272 SerValLeuProGlnLeuGluPheLysSerAsnAspAsnGluG 288
273 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3392 AATGGATGCTGAAGTTGATGTTTTTCTGAAGGAG..... 3427
288 uArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAsps 305
289 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3428 .....GAATGCCTGCCTTGGGCGATT 3449
305 erGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArg 321
306 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3450 CAGAAATTCCTA.....AAAAGCAGCTGAACAGTGCAGACTT...TTA 3490
322 PheAsnAspIleHis...ValProIleArgLeuGluCysValLysPheAl 337
323 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3491 GTCAGTGATATTCAGACAATTCAGCCCGAGTCAACAGTGTCAATGAAGG 3540
337 aserHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluT 354
338 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3541 TGGCAGAGAGATAAAGATGAAGCA..... 3565
354 yrLeuLysValArgSerHisAspProGluAlaIleArgHisAspVal 370
355 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3566 .....GAGCCAGAGTTTGTCTCGAGA..... 3586
371 IleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAs 387
372 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3587 .....CTTGAGAGAGAACTCAAGAACTTAACACTCAGTGGGA 3624
387 pHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgV 404
388 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3625 TCACATGTGCCACAGGTC.....TATG 3647
404 alArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
405 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3648 CCAGAAAGAGAGGCGCTTGAAGGA.....GGTTGGAGAACTGTAAAGC 3691
421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLy 437
422 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3692 CTCAG.....AAAGATCTATCAGAGATGCACCAATGGATGAC 3729
437 sasLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuL 454
438 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3730 ACAAGCT.....GAAGAAGAGT 3746
454 euValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThr 470
455 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3747 ATCTTGAGAGAGATTTT...GAATATAAACTCCAGATGAATTACAG... 3790
471 ThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAs 487
472 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3790 ..... 3790
487 nAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgH 504
488 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3791 .....AAAGCAGTTGAAGAGATG..... 3808
504 isGlnValLysAspLeuLeuAspLeuLysGlnProLysThrAspAla 520
505 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3809 .....AAGAGAGCTAAAGAGAGGCC 3829
521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPr 537
522 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3830 CAACAAAAGAA.....GCGAAAGTGAACCTCCTTACTGAGTCTGTAAA 3873
537 oAsp.....ProGlyLysAlaGlnAspPheMetLysLysP 549
538 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3874 TAGTGCTACGTCAAGCTCCACCTGTAGCACAA..... 3907
549 heThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluVal 565
550 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3908 .....GAGGCCTTAAAAAAGGAAGTGAAGT 3934

```



CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and  
 CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins  
 CC may also be used to prevent diseases such as neural disorders, immune  
 CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
 CC AAB54007 represent sequences used in the exemplification of the present  
 CC invention.

SQ Sequence 1939 BP; 483 A; 527 C; 685 G; 235 T; 9 other;

#### alignment\_scores:

Quality: 181.50 Length: 337  
 Ratio: 1.107 Gaps: 11  
 Percent Similarity: 48.665 Percent Identity: 21.958

#### alignment\_block:

US-09-512-581-2 x AAC98079 ..

Align seg 1/1 to: AAC98079 from: 1 to: 1939

```

1122 ProGlyLysProLysThrThrAsnValLeuGlyAlaVal..... 1134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 CCGGTGCGACCCACGCGTCCGCGGTAAACGCGGCACAGCTGCCAGCAG 105
1135 .....AsnLysProLeuSerSerAlaGlyLysG 1144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 GATGGAGAGCGACTCAGACTCAGACAAGAGTAGCGACAACAGTGGCGTGA 155
1144 InSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 AGAGGAAGACGCGCTCGGTAAAGATGTCTGCGAAGACGAGCCGCAAG 205
1161 SerSerAsn.....ProSerProGlyArgIleLysG 1172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 GCCTCCAGGAGCTGGATCAGCCAGCGTGTCCCATCGAGAGAGAGAA 255
1172 yArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThr. 1188
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 CTCGAAAGCTCATCTGACTGCGAAGAACACGACGACGAGGACTTCACAC 305
1189 .....MetSerSerProLeuProGly 1195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 CTGAGAAGAAACAGCGGTCCGGGCGCCACGAGGGGCGCTCTGGGGGA 355
1196 LysLys.....SerAspLysArgAspAs 1203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 CGGAAAGAAAGAGCGCGTCCGCTCCGACTCCGACTCCAGGCCGA 405
1203 pSerAspLeuValArgSerGlu..... 1210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 TTCGACGGGGCAAGCTGAGCGGTGCGCATGGCGCGTCCGCGTCT 455
1210 ..... 1210
456 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 505
1211 .....LeuGluLysProArgG 1216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 CCTCCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555
1216 yArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspA 1233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
556 GCGGAACCGAAGCTGAACGCGCTCCGTCAGCTCCAGCAGTGCAGAGT 605
1233 spluThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSer 1249
|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

606 ACAGCGACAGAGTGGACCCATCAGTGTGGAAGCGGGGACGAGCG 655
1250 ArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProG 1266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 CGAGCGCGGAGCTGGAGCCCGCGCGCGGAGAGCAGGAG...GAGGA 702
1266 uGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsn 1283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 GCTGGCGCGCTGGGAG.....CAGGAGAAGGAGGAGGAGC 743
1283 exProProLysLysGlyLysArgGlyArgProProLysProLeuGly 1299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 GGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
1300 GlyThrProLysGlu.....GluProThrMetLysThrSe 1311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
794 AGCAGCGGGGACGAGCTCAGGGAGGAGGAGGAGCGCGTCAAGAAGCG 843
1311 rLysLysGlySerLysLysSerGlyProProAlaProGluGluG 1328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
844 AGCAAGGCG.....CGGGCGGGGTCGCCCGTCTCTCTGACTCG 887
1328 luGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
888 AGCCGAGGCG..... 898
1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
899 .....GAGCTGGAGAGAGGCGCAAGAAATCAGCGAAGAGCGCGC 942
1361 rSerAlaIleGluSerThrGlnSerThrProGlnLys...GlyArgG 1377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
943 CTCAGCACAGAGCGCGCGCGAGAACTGCGCCAGAGAGAGAGAGTGC 992
1377 rgProSerLys 1380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
993 GGCCCCGAGGAG 1003

```

seq\_name: /SIS1/gcgdata/geneseq/geneseq/NA1989.DAT:AA90338

seq\_documentation\_block:

ID\_AA90338 standard; cDNA; 12923 BP.

XX AC AA90338;

XX DT 29-MAR-1992 (first entry)

XX DE Sequence of human muscular dystrophy (MD) cDNA.

XX DE Dystrophin; muscular dystrophy; probe; antibody; diagnosis;  
 KW prenatal; heterozygote; gene therapy; genetic screening;  
 KW foetal screening; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 209..12923  
 XX FT /\*tag= a

XX PN W08906286-A.

XX PD 13-JUL-1989.

XX PF 16-DEC-1988; 88WO-US04504.

XX PR 22-DEC-1987; 87US-0136618.

XX PA (CHIL-) CHILDRENS MED CENT.

XX PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;

XX DR WPI; 1989-220587/30.

XX DR P-PSDB; AAP90373.



1045 .....GluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056  
 4112 CTATGATTATATGATGATATAAAAAA...AATATAGTACGATGAA 4158  
 1057 AspAspAlaLysMetAsnGluLysLeuThrValCysAspValAlaMet 1073  
 4159 AATAATTATAAATTAAATGAAAGATG.....AACATATCTAT 4196  
 1073 t..... 1073  
 4197 GTCAATGATGAAGATATGATTCCTACGTTAAATAGTGAACATGGAATA 4246  
 1074 .....AsnIleIleMetSerLysSerThr..... 1081  
 4247 ATTTTCCAGTTGTCACCGCAATTATAGAAAAAAGTACTTATATA 4296  
 1082 .....ThrTyrSerLeuGluSerProLysAspProValLeuProAl 1095  
 4297 GATTGAACTTATATGATAGTAAATTCATGGCAGT..... 4332  
 1095 aArgPheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrL 1112  
 4333 .....TTTACAGAAGAAAAATATAATTTTGTAAATGAATGATT 4375  
 1112 eu.....ProProGluMetLysSerPhePheThrProGlyLysPro... 1125  
 4376 TATTCATACTAAAGATGGGAGTTTAAATTTTCCAGGGTAAAAATCTG 4425  
 1125 ..... 1125  
 4426 TTTAATAATAAATTTTAAATGATATCTAATCAGGATGGTGTTTCCTT 4475  
 1126 ....LysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSera 1141  
 4476 TTTTAAAAATATGAATCTTTTAGGGACCTTAATAAA..... 4512  
 1141 laGlyLysGlnSerGlnThrLysSerArgMetGluThrValSerAsn 1157  
 4513 .....TCAATATAGCTTAAACAGTAGAGAGTGTAAANAAT 4548  
 1158 AlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLe 1174  
 4549 AGTAATAATAATTTAGTAAATAT.....AAGGTGATGA 4583  
 1174 uAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSers 1191  
 4584 TAATATTGGAAATATGGAGAAATATCAATACAACAATGTTACAATTGCGA 4633  
 1191 erPro.....LeuProGlyLys.....LysSerAspLys 1200  
 4634 GTGATGAACATATATCTACAAGGAGAGATATACACGAGCAATCATTTCT 4683  
 1201 ArgAspAspSerAspLeuValArgSerGluLeuLysProArgGlyArg 1217  
 4684 AGACAGCATATGATGTTATCTTTTAAANAATTGAA...GCTAGAAGTAA 4730  
 1217 gLysLysThrProValThr...GluGlnGluLysLeuGlyMetAspA 1233  
 4731 AAAATATAGTATACCTTATATAATAGGACAAAAAGTAATTTGGAGA 4780  
 1233 spLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgSer 1249  
 4781 ATGACATGAGACTATTAATGAATATGAA..... 4809  
 1250 ArgLysArgGlyHisThrAlaSerGluSerAspGluGlnInTrpProGl 1266  
 4810 .....AATGTATGTAGTAACATAGATGTTAATGAATGGGAAGA 4847  
 1266 uGlu.....LysArgLeuLysGluAspIleLeuGluAsnG 1278  
 4848 TAAGGTAATGGTACATGTAATAGTGTGGTGATAAGAGAGACTGAAAAAGA 4897

1278 luAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPro 1294  
 4898 ATATGAAGAATAATGAAAGAATAATGAAAGAAT..... 4935  
 1295 LysProLeuGlyGlyThrProLysGluGluProThrMetLysThrSe 1311  
 4936 .....AATGAAAGAATAATGAAAGAATAA 4961  
 1311 rLysLysGlySerLysLysSerGlyProProAlaProGluGluGluG 1328  
 4962 TGAAGAATAATGAAAGAATAAT.....GAAAAAGA 4993  
 1328 luGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLys 1344  
 4994 ATATGAAGAAATAATGAAAGAATAATGAAAGAAATAATGAAAGAAAT 5043  
 1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSe 1361  
 5044 AATGAAGAA.....AATAATGAAGAAAA 5066  
 1361 rSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgP 1378  
 5067 TAATGATATAGAAAGAATGAT..... 5088  
 1378 roSerLysThrProSerProSerGlnProLysLysAsnVal 1391  
 5089 ..ATAAGGATAATAATTCGGGCAAGTGAAGAAATAATATA 5127  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98079  
 seq\_documentation\_block:  
 ID AAC98079 standard; cDNA; 1939 BP.  
 XX  
 AC AAC98079;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:89.  
 XX  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antinefactive; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055351-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05883.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-587534/55.  
 DR P-PSDB; AAB53322.  
 XX  
 PT Colon cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer -  
 XX  
 PS Claim 1; Page 532; 2104pp; English.  
 XX  
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;







211 nLysGlnAla.....TyrAspLeuAlaLysAlaLeuLeuLysArgThrA 226  
||||| : : : |||| : : : : : : : : : : : : : : :  
2007 TAAGAAAGAACCCTTTTTCGTCTAAAAATAATTCACTAAGCACAATC 2056  
:  
226 laGlnAlaIleGluProTyrIleThrThr.....PhePheAsnGln... 239  
: : : ||||: : ||||: : ||||: : ||||: : ||||: :  
2057 CAAAAGAAAATAATATATATATACATCTCGGAATCTTTTATCATGTG 2106  
:  
240 .....ValLeuMetLeuGlyLysThrSerIleSerAs 250  
|||||: : : ||| : : : : : : : : : : : : : : :  
2107 CAGGATAAGGAAAAGACTGTTCTACTTAAAAAAGAAAGAAATAATGA 2156  
:  
250 pLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsps 267  
| :  
2157 T.....AAGAAATACATTAGCTCTCTGTTTAATAAACCAATAATATACAA 2200  
:  
267 erHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSer 283  
: : : : : ||| : : : : : : : : : : : : : : :  
2201 CATATACATTACAAAATGGAGTA..... 2223  
:  
284 AsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPh 300  
||| ||| :  
2224 AATRAAAAT.....TTAAATATGTT 2243  
:  
300 eGlyAlaLysaspSer.....GluLeuAlaSerGlnAsnLysProLeuT 315  
:||| : : : ||||| :  
2244 AGGAATAAGACATTCTATTTATAAATAGATGAAAGAACCAATATGTTGA 2293  
:  
315 rpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeu 331  
: : : : : ||||| : : : ||||| : : : : : : : : : : : : : : :  
2294 AAGAAATGTTAATATGGAAAT....AATCATAGTATAATAAGAAAAAAG 2340  
:  
332 GluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAl 348  
: : : ||||: : : : : ||| : : : : : : : : : : : : : : :  
2341 AAGAAAAAAAATTTATCTTTTCATGT.....GATATTAT 2375  
:  
348 alYasap.....LeuThrGluTyrIleLysValArgSerHisAspProGluG 364  
: : : : : ||||| : : : ||| : : : : : : : : : : : : : : :  
2376 AAATGATAATATTACACCTTATGAATCAGATAAGAGAAAAACAATTCTA 2425  
:  
364 luAlaIleArgHisaspvalIleValserIleValThrAlaAlaLysLys 380  
: : ||||: : : : : ||||| : : : : : : : : : : : : : : :  
2426 ATAATATTAAAGAGT.....ATGGATATATTAAATATGTATGAAAAAGA 2466  
:  
381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgTh 397  
||| ||||: : : ||||: : : : : ||||: : : : :  
2467 AAACCAACCTTTATATAAT.....TTATCTCGAACAGGATCTCTAC 2510  
:  
397 rLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaG 414  
| : : | :  
2511 TGTAGATGCATAATAAAATAATAATAGTGAAGATAATATAAATATACAGA 2560  
:  
414 InIleTyLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAla 430  
: : ||| ||| ||| ||| : : : : : : : : : : : : : : :  
2561 GAACAAATAAAATATATGAATTG..... 2583  
:  
431 LysGlnIleAlaTpIleLysAspLysLeuLeuHisIleTyrTyrGlnAs 447  
: : : : : ||||: : : : : : : : : : : : : : :  
2584 .....AGTAAATAAGAAATTAGAAATTTAAATTTGTA 2615  
:  
447 nSerIleAspArgLeuLeuValGluArgIlePheAlaGlnTyrMet.. 463  
|||: : : ||||: : : : : ||||: : : : : : : : : : : : : : :  
2616 TAGTATGCATGAATATTTTAAGGTGCTCTTAAGGAAAAAATAATATAG 2665  
:  
464 .....ValProHisAsnLeuGlu.....Thr 470  
: : : : : ||||: : : : : : : : : : : : : : :  
2666 ATATATATTTCTAAATAATATGGAAGAGTAACATATAAAAAATGAAATGATA 2715  
:  
471 ThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAs 487  
:  
2716 AATGAAAAGATAAGTAGATGGAGATATATATATATCTCTGTGCACAAAA 2765  
:  
487 nAlaValLysAlaLeuAsnGluMetTrpLysCys.....G 499



```
1079 .....LysSerThrThrTyrSerLeuGluSerProLysAsp 1090
      ::::: ::::: ::::: :::::
9429 ACCGAAACACATCGAAACCTTTGATCAGAACATAGAACAAATCACAAG 9380
      ::::: ::::: ::::: :::::
1091 ProValLeuProAla...ArgPhePheThrGlnProAspLysAsnPhese 1106
      ::::: ::::: ::::: :::::
9379 TGGATCATTCATCGAGATGAACTTTTAGATGAGTCTGAAAGAAAGAAACC 9330
      ::::: ::::: ::::: :::::
1106 rAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrProG 1123
      ::::: ::::: ::::: :::::
9329 ACAACAAAGGAGACATCTCTAAACGCTTAAAGGCTGAATGAATGACA 9280
      ::::: ::::: ::::: :::::
1123 LyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSer 1139
      ::::: ::::: ::::: :::::
9279 TCGGCCAAAGGTGACTCCACACGTCACCAAGCAGCAAAATTTGATGGCA 9230
      ::::: ::::: ::::: :::::
1140 SerAlaGly.....LysGlnSerGlnThrly 1148
      ::::: ::::: ::::: :::::
9229 AACCGGGTGACCACTGCAGGAAAGTAGAGAGCCCAAAATCTCTGAGCT 9180
      ::::: ::::: ::::: :::::
1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerAsnPros 1165
      ::::: ::::: ::::: :::::
9179 CAACCGTCGATTGCGAGTATTCTCACAGAAATTAAGACTGCAAGGCCT 9130
      ::::: ::::: ::::: :::::
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHis 1181
      ::::: ::::: ::::: :::::
9129 CCATTCTCT.....TTGAAGGAATTGGAGCAG 9104
      ::::: ::::: ::::: :::::
1182 SerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLys..... 1196
      ::::: ::::: ::::: :::::
9103 TTTAACTCAGATATACAAAATTTGCTTGAACCACTGGAGGCTGAATTC 9054
      ::::: ::::: ::::: :::::
1197 .....LysSerAspLysArgAspA 1203
      ::::: ::::: ::::: :::::
9053 CGAGGGGTGAATCTCAAGAGCAAGACACTTCAATAAGATATGAGTGAAG 9004
      ::::: ::::: ::::: :::::
1203 spSerAspLeuValArgSerGluLeuGluProArgGlyArgLysLys 1219
      ::::: ::::: ::::: :::::
9003 ACAATGAGGTACTGTAAATGAATTGTTCAAGAGAGGAGACAACTTACAA 8954
      ::::: ::::: ::::: :::::
1220 ThrProValThrGluGluGluLysLeuGlyMetAspLeuThrly 1236
      ::::: ::::: ::::: :::::
8953 CAAGAATCACAGATGAGAGAGGAGGAGAAATAAGATAAAACAGCA 8904
      ::::: ::::: ::::: :::::
1236 sLeuValGlnGluGln.....LysProLysGlySerGlnArgS 1249
      ::::: ::::: ::::: :::::
8903 GCTGTTACAGACAAACATAATGCTCTCAAGGATTTGAGGTCTCAAGAA 8854
      ::::: ::::: ::::: :::::
1249 erArgLysArgGlyHsThrAlaSerGluSerAspGluGlnThrPro 1265
      ::::: ::::: ::::: :::::
8853 GAAAAAAG.....GCCCTAGAAATTTCTCACCAGTGGTAT 8819
      ::::: ::::: ::::: :::::
1266 GluGluLysArgLeuLysLysAspIleLeuGluAsnGluAspGluGlnAs 1282
      ::::: ::::: ::::: :::::
8818 CAGTACAGAGCAGCGCTGATGATCTCTGAATGCTTGGATGAATTT... 8771
      ::::: ::::: ::::: :::::
1282 nSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyG 1299
      ::::: ::::: ::::: :::::
8771 ..... 8771
1299 LyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySer 1315
      ::::: ::::: ::::: :::::
8770 .....GAA 8768
      ::::: ::::: ::::: :::::
1316 LysLysLysSerGlyProProAlaProGluGluGlu.....GluG 1329
      ::::: ::::: ::::: :::::
8767 AAAAATTAGCCGCTACTGACCCAGAGATGAAGAAAATTAAGGA 8718
      ::::: ::::: ::::: :::::
1329 uGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnH 1346
      ::::: ::::: ::::: :::::
8717 AATTGATCTGTAATG.....CAGAAAGAAAGAGGAGCTGA 8680
```

```
1346 IsArgValSerArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
      ::::: ::::: ::::: :::::
8679 ATGCAGTGCAGGAGCAAGCTGAGGCTTCTCTGAGAAATGGGCGCGCAATG 8630
      ::::: ::::: ::::: :::::
1363 AlalleGluSerThrGln 1368
      ::::: ::::: ::::: :::::
8629 GCAGTGGAGCCCAACTCAG 8612
      ::::: ::::: ::::: :::::
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA70106
seq_documentation_block:
ID_ AAA70106 standard; DNA; 7458 BP.
XX AAA70106;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:239.
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX Plasmodium falciparum.
XX W0200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX Disclosure; Page 462-464; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX Sequence 7458 BP; 3523 A; 514 C; 1174 G; 2247 T; 0 other;
XX alignment_scores:
```



```

11150 CAGTTTAAATAGT.....GTTAATGAAGTGGGCAGAGATAA 11113
602 yspHeLeuLeuGluAArgIleAlaProValHisIleAspThrGluSerIle 618
11112 AGAGTGAAGCTGAACGTGAGTTGGATCCAGACTGGAGACAGAA..... 11069
619 SerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAs 635
11068 .....CTTAGAGAGCTTAACACTCAGTGGGATCACAATATGC..... 11033
635 pGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGlu. 651
11032 ....CGCCAGGTCTACACCAAGAAAGGAGCTTAAAGCGAGGTTGGATA 10987
652 .....LeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
10986 AACCGTAAAGCCCTCCAAAAGATCTATCAGAGATGCATGAGTGGATGACA 10937
665 HisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAs 681
10936 CAGCTGAAGAAGATATCTAGAGAGATTTTGAATATAAACTCCAGA 10887
681 pGlu.....LysValAlaGluAlaAla 689
10886 TGAATTACAGACTGCTGTTGAAGAAATGAAGAGAGCTAAAGAAGAGGCAC 10837
689 euGln.....IlePheLysAsnThrGlySerLys 698
10836 TACAAAAGAAACTAAAGTGAACCTCCTTACTGAGACTGTAATAATAGTGA 10787
699 IleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHi 715
10786 ATAGCTCAGCTCCACCTCAGCAGACAGAGGCTTAAAAAGGAAGTGA 10737
715 sHisLysSerLys.....LysGlyProP 723
10736 AACTCTGACCACCAACTACCAATGGCTGTGCACCAGGCTGAATGGAAT 10687
723 roArgGlnAlaLysTyraIleHisCysIleHisAlaIlePheSerSer 739
10686 GCAAACTTTGGGAAGATTGGGCATGTTGGCATGAGTTATTGTCTAT 10637
740 LysGlu..... 741
10636 TTAGAGAAACCAACAAGTGGCTCAATGAAGTAGAATTGAAACTTAAAC 10587
742 .....ThrGlnPheAlaGlnIlePheGluP 750
10586 CATGGAAATGTTCTCTGAGGACCTGAGGAAATCACTGAAGTGTAGAT 10537
750 roLeu.....HisLysSerLeuAspProSerAsnLeuGlu 761
10536 CTCTTGAAATCTGATGCATCATTCAGTACAGAGAGAACCCAAATCAGATTCT 10487
762 HisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778
10486 CTATTGGCACAGACTTTACAGATGAGGAGTCTGATGAACTGATCAA 10437
778 oAspGln.....PheAlaAlaProThrLysSerTrpValAlaThrP 792
10436 TGAGGAGCTTGACAGCTTTAATCTCGTTGGAGGGAACATACATGAAGAG 10387
792 heIleValLysAspLeuLeuMetAsnArgLeuProGlyLysLysThr 808
10386 CTGTGAGGAAACAAAAGTTGCTTGAACACAGATATCCAGTCTGCCAGGAA 10337
809 ThrLys.....LeuTrpValProAspG 816
10336 ATTGAAAGTCCCTTGCACTTAATTCAGGATCCCTTGAATTCATTGACAA 10287
816 uGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetV 833

```

```

10286 GCAGTTGGCAGCTTATATCACTGACAAGTGGATGCAGCTCAAATGCCTC 10237
833 alArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849
10236 AGGAA.....GCCAGAAAATCCAATCAGATTTGACAAGTCAT 10199
850 ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGl 866
10198 GAGATTAAGTTTAGAAGAAATCAAGAACATACACGGGGAAGGATGCCAA 10149
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaG 883
10148 CCAAGGGTCTTTTCAAAATTCATGTTGCACAGAAAAAATTTACAAGATG 10099
883 lySerAlaIleValLysLeuAlaGlnGluProCysTyThrHisGlu..... 897
10098 TCTCCATGAATTTTCGATTATCCAAAACACGCCAATTTTGAACAACGT 10049
898 .....IleIleThrLeuGluGlnTyThrGlnLeuCysAl 908
10048 CTAGAGAAAGTAAGATGATTTTAGATGAAGTCAAGATGCTTTTGCCTGC 9999
908 aLeu.....AlaIleAsnAspGluCysTyThrGlnValArgGlnValP 922
9998 ATTGGAAACCAACAGAGTCTTGAACAGGAAGTAATTCAGTCACAACCTAAGTC 9949
922 heAlaGlnLysLeuHisLysGlyLeuSerArgLeuArg..... 934
9948 ATTGTGTGAACTTGTATAAAAGCCTGAGTGAAGTCAAGTCTGAAGTGGAA 9899
934 ..... 934
9898 ATGCTGATTAACACCGGACGCTCAAAATGTACAAAAAGCAGACAGAAAA 9849
935 .....LeuProLeuGluTyThrMeta 941
9848 TCCCAAAGAGCTTGTCAACGAGTAAACAGCTTTTGAATTCATTAACAATG 9799
941 laIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
9798 AGTTGGGTGCG.....AAGGTAACAGAGAGAGAAAGCAACAG 9764
958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyThrLeuLy 974
9763 TTGAGAAATGCTTGAAGTGTCCCGTAGATGAGAAAGGAA..... 9722
974 sGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrv 991
9721 .....ATGAATGTCTTAACAGAAATGG. 9701
991 alValProTyThrIleHisLeuLeuAlaHisAspProAspTyThrValLys 1007
9700 .....CTGGCAGCACACATACAGAATTCAGCAAG 9671
1008 ValGlnAspIleGluGlnLeuLysAsp.....ValLysGluCysLeuTr 1022
9670 AGATCAGAGTTGAAGAAATGCCAAGTAATTTGGATTCTCGAAGTTGCCTG 9621
1022 pPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlap 1039
9620 G.....GGAAAGGCTACTCAAAAAGAGATTGAGAAACAGAAAGGCTC 9580
1039 heIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
9579 ACTTGAAGAGTGTACAGAAATTAGAGAGAGTCTTTGAAAATGGTGTGGGC 9530
1056 ProAspAlaLysMetAsnGluLysLeuTyThrVal..... 1068
9529 AAGAAGAAACCTTGGTAGAAGATAAAGTCTTCTGAAACAGTAAGTCTG 9480
1069 .....CysAspValAlaMetAsnIleMetSer. 1078
9479 GATAGCTGTACCTCCAGAGTAGAAGAAATGGCTAAATCTTTTGTGGAAT 9430

```







```

2842 GGCTTCTCATCAGTTGAAATGGAACTCAATTCA.....A 2876
890 lacInGluProCysTyrHisGluIleThrLeuGluGlnTyrGlnLeu 906
2877 TCGAGGAACAACATTACAGGCCAGCAATCTTTAAACAAATGAAAA 2926
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
2927 GAAGACCAACAACCTTCAGGGGAACATAAATGAGCTAAACCAATCAAGTCA 2976
923 agInLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrM 940
2977 ACAGAAG...AAAAACAATAATTGAAGCACTCCAAGGAGAGCTTAAATTC 3023
940 etAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
3024 CAGTTTACAGAGACAGAGCTTGAGAATAAATACAGCAGCAGCTTAACA 3073
957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgGluTyrLe 973
3074 CAGGCAGCCGGAACCTT..... 3091
973 uLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGluT 990
3092 .....GCAGCAGAGAAGAAAGAAATATCAGTATTACAAAACAAC 3131
990 yrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
3132 ATGAAAAAGTCAGGAAATTTCAACACAGCTTCAATCTGATTTCTATGGG 3181
1007 LysValGlnAspIle.....GluGlnLeuLysAspValLysGln 1019
3182 AGGGAATCTGAACCTTCTACCCAGCAGCAAGATCTTAAGTCTGTAGAACA 3231
1019 uCysLeuTrpPheValLeuGluLeuLeuMetAlaLys..... 1031
3232 GAAGCTTTCTAGCACAGCAGGAGCTTGATTTCAACAGAAATCAAATTC 3281
1032 ..AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
3282 GAATCAAAAT.....AAATGATTCAAGAACTG 3310
1048 LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
3311 AAGACTGCCAAGGCTACATTTGGAGCAGGATTCAGCAAGAAGAAACAGCA 3360
1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerT 1081
3361 ATTGCAGGAGCGATGTAAAGCACTACAAGACATTCAGAAGAAAGTCAC 3410
1081 hrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097
3411 TG..... 3412
1098 PheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPr 1114
3413 .....AAGAAAAAGAACTGGTAATGAGAGACTTAATTC...GC 3450
1114 oGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnVal 1131
3451 AGAGATAGAAGAAAT.....AAATGTAGACAGAAAAAGAAA 3488
1131 euGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr 1147
3489 TCATAAATCAACGAAAGAACTC..... 3511
1148 LysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsnPr 1164
3512 AATCCCAACAACTAGAACGATA..... 3535
1164 oSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
3535 ..... 3535

```

```

1181 issSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLys 1197
3535 ..... 3535
1198 SerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPr 1214
3536 .....AAGGAGATAACAAATCT 3552
1214 oArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyM 1231
3553 TAAAGATGCTAAACAGACTTCTAATTCAGCAGAAATTAGAACTTCAAGGAA 3602
1231 etAspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGln 1247
3603 AAGCGACTCCCTGAAGCAGCTGTGAACAGAGAGAGAGAAATCAGCAG 3652
1248 ArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
3653 ATACTAAAA.....GACCAGGTGAA 3672
1264 pProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluG 1281
3673 AAGCAAGAAGAGGAGCTGAAGAAGAAATTTATTGAGAAGAGCTTAAGT 3722
1281 lAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeu 1297
3723 TCATTTCC..... 3730
1298 GlyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysG 1314
3731 .....GAATAAAAGAAAGAACTAGG 3753
1314 ySerLysLysLysSerGlyProProAlaProGluGluGluGluGluG 1331
3754 AATGAAGAAG.....CATGAAGAAATGAGG 3779
1331 luArgGlnSer.....GlyAsnThr 1337
3780 CTAACTTACCATGCAGATTACAGATTAAATGAAACTTAGGCCTGTG 3829
1338 GluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnG 1354
3830 AAGAAGGAGTGGCAATCTAGTCAACGAGAGTTAGTGAGTTGAGAAACA 3879
1354 nArgAlaGlu...SerProGluSerSerAlaIleGluSerThrGlnSerT 1370
3880 AACGGATGACTTACGGGGTGAAATTCAGTATTAGAAGCAACGGTTTCA 3929
1370 hrProGln.....LysGlyArg 1375
3930 ATAACTCAAGATGAAGGAGAGCACTACTGAAAGATGCTTTAAAGGAGAA 3979
1376 GlyArgProSerLys 1380
3980 GTGGAATAGAAAAG 3994
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AA27558
seq_documentation_block:
ID AA27558 standard; cDNA; 19307 BP.
XX
AC AA27558;
XX
DT 07-AUG-1996 (first entry)
XX
DE Shuttle vector pAdDel.CMVdys.
XX
KW Adenovirus type 5; Ad5; vector; gene therapy; gene transfer;
KW helper virus; dystrophin; muscular dystrophy; ds; cyclic.
XX
OS Synthetic.
XX

```



```

296 uAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysP 313
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1558 TCAGCAAGACACAGCGCAAACTTCGAGAA.....GCTCAGAAT.... 1597
313 roLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle 329
1597 ..... 1597
330 ArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAs 346
1597 ..... 1597
346 pLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProG 363
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1598 .....GATTTGGACAAGTTCCTAGCT...CAAAATTGGCGATAAGG 1634
363 luGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLys 379
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1635 ACCAAAGATCCGAGAC.....CTTGAAGCTTTATTACAGAAGAGTAAA 1678
380 LysAspIleLeuLeuValAsnAspHis..... 388
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1679 GAAATATTTCATTACTAGAAAAAGAGAGAGATCTTTATGCAAAAAT 1728
389 .....LeuLeuAsnPheValArgGluArg. 396
1729 TCAGCGTGTGAAGGAGACACTGCTGTTCTTAACCAAGTTACAGAAAAA 1778
397 .....ThrLeuAspLysArgTTPArgValArgLysCluAlaMetMetGly 411
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1779 ACCATACACTACAGGAGCAA..... 1798
412 LeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAs 428
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1799 GTAACCTCACTAACAGAGAGCTGAAGATCAGTCAGAAAGTCATAAACA 1848
428 pAlaAlaLysGlnIle...AlaTrpIleLysAspLysLeuLeuHisIleT 444
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1849 AGCCGAGGAGAAATTTGCATGACCAGGTACAGAGCAGAGGACACATCTT. 1897
444 yTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1898 .....AGACTGCACAGACCGTGCTT..... 1921
461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLe 477
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1922 .....TCCCTAGAAACTAGTGTCAAT..... 1942
477 uTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnG 494
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1943 .....GAATTAATAGT.....CAATTAATAG 1964
494 luMetTrpLysCysGlnAsnLeuLeuLeuArgHisGlnValLysAspLeu 510
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
1965 AAAGCAAGAGAGAGTCTCCAGCTTGACATACAGATTAAAGCCAAACC 2014
511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSe 527
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2015 GAACATTATTACTACAGCAAGCAGCAAAAACTGCTCAAGAGCTGATCT 2064
527 rLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnA 544
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2065 TCAGAAATCATTTGGACACAGCTCAAAATGCATTACAAGATAAACACAGC 2114
544 spPheMetLysLysPheThrGlnValLeuGluAspGluLysIleArg 560
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2115 AGTTAAATAGATTACTACTAGTTGGATCAGTCAAGTCAAGTTACAA 2164
561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGl 577
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2165 GACAAGCAAGAA.....CATTCGAGTCAGCTGGA 2193
577 uGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProT 594

```

```

|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2194 AAGTCATCTTAAAGAAATAAAGAGAAATACCTCTCTTTAGAACACAAAA 2243
594 hrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPro 610
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2244 CCGAAGAGCTAGAAGGTCAATTAAGAACTA..... 2275
611 ValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLy 627
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2276 ...GAAGCTGATAGTCTTGAAGTTAAAGCAAGCAAGGAGCAGGCTTTGCA 2322
627 sSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspG 644
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2323 AGATCTA.....CAACAGCAAAAGACAGCTGAACACAGATT 2357
644 lnaIleAlaArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2358 TAGAGCTCAGAGCC...ACAGAATTGAGTAAACAACCTT..... 2392
661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLe 677
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2393 .....GAAATGGAGAAGAAATAGTATCCAGTAC 2421
677 uLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysA 694
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2422 AAGATTGGATCTACAGAAAAATCTGAA...GCCCTTGAAAGTATCAACG 2468
694 snThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaLeu 710
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2469 AAAAGCTTACCAGCAAGCAAGGAGAAAAACAATCTGAAACAAGATTTT 2518
711 LeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLy 727
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2519 GAACTTTTA..... 2527
727 sTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnP 744
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2528 .....AGTCAAGAAACAAAGA 2543
744 heAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2544 TTCAG.....CATGAGGAATTGAAT...AACAGAATT 2572
761 GluHisLeuIleThrProLeu.....ValThrIleGlyHisIleAlaLe 775
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2573 CAACAACAGATAACAGAACTACAAAAAGTGAATGGAGAAAGAGCTTT 2622
775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrP 792
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2623 AATGACAGAG.....CTTCTTACA. 2641
792 heIleValLysAspLeuLeu.....MetAsnAspArgLeuProGlyLys 806
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2642 .....GTAAGGACAACTATCAAAAGTTCTGATTCTTTGAAAAACTCT 2686
807 LysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMe 823
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2687 AAAAGT.....GAATTTGAAAAGGAGGAATCA 2712
823 tValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetL 840
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2713 GAAAGGAAAGCCGCTATATTAGACTTGGAAAAAACTTCCAAGAAATTAA 2762
840 ysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2763 AGCATCAACTTCAAGTCAGAGTGAACACACACTTAAG..... 2800
857 lIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysPr 873
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
2801 .....GAACAGAGGAAGTCACTGAAAAAGTCACTTGAAGAAAGGA 2841
873 oAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |

```











[illegible]



1366 TGCTCATGAGGATTGATGATGATCTGACATCTCATCAAGGACTTGTG 1415  
292 .....ValValLysLeuLeuAlaLysMetPheGly.....Ala 302  
|||:||||| |||:|||||  
1416 GTAATGTTCTACAGTTAGGAAGTCAACTAGTTAGTGGAAAGGGAATATCA 1465  
303 LysAspSerGluLeuAlaSerGlnAsnLys.....ProLe 314  
|||:||||| |||:|||||  
1466 GAAGATGAAGAAGCTCAAGTGAAGAACAATAATCTCTCTAAATTCAG 1515  
314 uTrpGlnCysTyr.....LeuGlyArgPheAsnA 324  
|||:||||| |||:|||||  
1516 ATGGGAATGTCACGGGTAGCTAGCATGGAAAAACAAGCAATAATACACA 1565  
324 spileHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340  
|||:||||| |||:|||||  
1566 AAGTCTAATGATCTCCAGATCCAGAAATTAAGAAGAACTAGATGACTGG 1615  
341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu..... 353  
|||:||||| |||:|||||  
1616 TTAACAAAACCTGAAGAGAGAAGCTAAGAAATGGAGGAAGCCCTTTGG 1665  
354 .....TyrLeuLysValArgSerHisAsp..... 361  
1666 ACCTGATCTTGAAGATCTAAAAATGCCAAGTACAACAACATAACGCTTTC 1715  
362 .....ProGluGluAlaIleArg.....Hisasp 369  
|||:||||| |||:|||||  
1716 AAGAGATCTAGAACAGGAGCAGGTCAGCGTCACTCGCTCACTACATG 1765  
370 ValIleVal.....SerIleValThAlaAlaLys 379  
|||:||||| |||:|||||  
1766 GTAGTAGTGTGTTGATCAATCCAGCGGTGATCAACACAGCTGCTTTGGA 1815  
379 slyAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluA 396  
|||:||||| |||:|||||  
1816 AGAACAACTTAAGGTACTGGGAGATCGAGGGGCAATATCTGC.....A 1859  
396 rgThrLeuAspLysArgTTPArgValArgLysGluAlaMetMetGlyLeu 412  
|||:||||| |||:|||||  
1860 GATGGACTGAACACCGCTGGATGTTTACAGATATCTTCTAAATGG 1909  
413 AlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAl 429  
|||:||||| |||:|||||  
1910 CAGCATTTTACTGAAGAAGCAGTGCCTTTTACT..... 1942  
429 aAlaLysGlnIleAlaTTPileLysAspLys.....LeuLeuHisI 443  
|||:||||| |||:|||||  
1943 .....ACATGGCTTTCAGAAAAAGAGATGCAATGAAGAACA 1979  
443 lefTyrTyrGlnAsnSerIleAspAspArgLeuValGluArgIlePhe 459  
|||:||||| |||:|||||  
1980 TTCAGACAAGTGGCTTAAAGATCAAAATGAATGATGTCAGTCTTCAC 2029  
460 AlaGlnTyrMetValProHisAsnLeuGlu.....ThrThrGl 472  
|||:||||| |||:|||||  
2030 AAAATATCTACTTTAAAAATAGATCTAGAAAAAGAAAGCAACCATGGA 2079  
472 uArgMetLysCysLeuTyrTyr.....LeuTyrAlaThrLeuAspLeuAsnA 488  
|||:||||| |||:|||||  
2080 AAAACTAAGTTCACCTCAATCAAGATCTACTTTCGGCAGCTGAAAAATAAGT 2129  
488 laValLysAlaLeuAsnGlnMetTTP.....LysCys 498  
|||:||||| |||:|||||  
2130 CAGTGACTCAAAAGATGGAATCTGGATGGAAAACTTTCACACACGTTGG 2179  
499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLys... 514  
|||:||||| |||:|||||  
2180 GACAATTTAACCCCAAAAAAGAAAGAGTTTCAGCACAAAATTTTCACAGGC 2229  
515 .....GlnProLys.....ThrAspAlaSerValLysAlaIleP 526  
|||:||||| |||:|||||

2230 TGTCACCACTCAACCATCCCTAACACAGACAACTGTA.....ATGG 2273  
526 heSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542  
|||:||||| |||:|||||  
2274 AAACGGTAATGTTGACCAACAAGGAAACAATCATGTGTAACACATGCC 2323  
543 GlnAspPheMet.....LysLysPheThrGlnValLeuGl 554  
|||:||||| |||:|||||  
2324 CAAGACGAACCTTCCACCACCTCTCTCAAAAAGAGAGCAGATAACTGT 2373  
554 uAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrC 571  
|||:||||| |||:|||||  
2374 GATTCTGAA...CTCAGGAAAAGGTTGGATGTCGATATAACTGAACTTC 2420  
571 ysSerCysLys..... 574  
|||:|||||  
2421 ACAGTTGCATTACTCTCTCAGAAGCTGTATTACAGAGTTCTGAATTTCCA 2470  
575 .....GlnAlaGluCysValArgGluIleThrLysLysLeuGlyAs 589  
|||:||||| |||:|||||  
2471 GTCTAACGAAAAAGGCAACATCTCAGACTTGCAGAAAAAGTCAATGC 2520  
589 n...ProLysGlnProThrAsnProPheLeuGluMet..... 600  
|||:||||| |||:|||||  
2521 CATAGCAGGAAAAAGCAGAGAAAGTTCAAAAACTGCAAGATGCCAGCA 2570  
601 .....IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615  
|||:||||| |||:|||||  
2571 GATCAGCTCAGGCCCTGGTGGAAACACATGCAAAATGAGGGTGTATATGCT 2620  
616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh 632  
|||:||||| |||:|||||  
2621 GAAAGTATCATCAAGCTTCAAGAACAACTGAAC..... 2653  
632 rAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaG 649  
2654 .....ACGCGGT 2660  
649 lyLeuGluLeuLysValLeuSer..... 657  
|||:||||| |||:|||||  
2661 GCACAGAATCTGCCAATGCTGAGTGACAGAGACTTAACCTGGCTAGAGTAT 2710  
658 PheThrHisProIleSerPheHisSer..... 666  
|||:||||| |||:|||||  
2711 CAACCAACATCATTAACCTTTTATAATCAGCTACACAATTTGGAACAGAT 2760  
667 AlaGluThrPheGluSerLeuLeu.....AlaCysLeuLysMetAspA 681  
|||:||||| |||:|||||  
2761 GACAACTACTGCCGAAACTTGTGAAAAACCCAGTCTACCACCTATCAG 2810  
681 spGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySer 697  
|||:||||| |||:|||||  
2811 AGCCAAACAGCAATTAACCCAGTTAAAAATTTGTAAGGATGAAGTCAAC 2860  
698 LysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValle 714  
|||:||||| |||:|||||  
2861 AGATTGTACGCTCTTCAAGCTCAAAATTTGACAAATTAATAATTCAGAGCT 2910  
714 uHisHisLysSerLys...LysGlyPro.....ProArgGlnA 726  
|||:||||| |||:|||||  
2911 ACAACTGAAGAAAGGAGCAGGGCCCAATGTTCTGGATGCAGACTTGG 2960  
726 laLysTyrAlaIleHisCysIleHisAlaIlePheSer.....Ser 739  
|||:||||| |||:|||||  
2961 TGCCCTTTTACTAATCATTTTAACCAC...ATCTTGTGTTGTGAGGGCC 3007  
740 LysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAs 756  
|||:||||| |||:|||||  
3008 AAAGAGAAAGAGCTACAGACAATTTTGAC.....ACTTTACC 3045  
756 pProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773  
|||:||||| |||:|||||  
3046 ACCAATGGCTATCAGGAGACAATGAGTAGCATC..... 3079







```

823 MetValLysIleGlnAlaIleLysMetMetValArg..... 834
:::|||||::: ||| |||
2725 CCAATCAAGATTCAAGCCGCTCGTGAATGAAGAGAAATTCGAAATCGT 2774

835 .Trp.LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
::: |||:::|||||::: ||| |||
2775 CTACAAGGTTCGTTCCAGAACACGAGAAATCGATCATCTCCCTCACCA 2824
||| |||:::|||||::: ||| |||
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluG1 867
||| |||:::|||||::: ||| |||
2825 CTCGCCAGTTCCTTCCTTCGTTCCAGGAT..... 2857

867 nGlyIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884
::: |||:::|||||::: ||| |||
2858 .....TCTCAAGTT 2867

884 exAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
||| |||:::|||||::: ||| |||
2868 CGAATACATTGAGGCTGAAGAGAACCG..... 2896

901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG1 917
..... 2896

917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
::: |||:::|||||::: ||| |||
2897 .....TCGTTGTTCCACAATGCCAACAAGACTCA 2927

934 rgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysaspPro 950
GG..... 2929

951 ValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
||| |||:::|||||::: ||| |||
2930 ....AAATTCAGAAGTCTCAACTCTCTCGGACTTCAA.....GTCTC 2969

967 nValArgGluTyrLeuLysGlnHisAla.....A 978
::: |||:::|||||::: ||| |||
2970 CACAGCGGAAACATCTCAACCAACACACTCTGAAAACCTGGCTCTTG 3019

978 IValSerGluLysLeuLeuSerLeuLeuProGluTyr...ValValPro 993
||| |||:::|||||::: ||| |||
3020 CCGAGCAAGACTTCGAGGTTCCGTTGAGAACAGTACCTCCAGCCGAA 3069

994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
::: |||:::|||||::: ||| |||
3070 TTCACGCCCGCTCACCGTTGGACAACCTCGAGAGACCGAGCTCTCCA 3119

1010 PileGluGlnLeuLysAspValLysGluCysLeuThrPheValLeuGluI 1027
::: |||:::|||||::: ||| |||
3120 AATCAAGTACAACAAGATCTTCGAAAG.....GAGTTCGAGCTCGAA. 3162

1027 leuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMet 1043
||| |||:::|||||::: ||| |||
3163 .....CAAGAGACACCGAGAGCCGCGGTGAGTACTTCCACAGATG 3204

1044 ValGluAsnIleLysGlnThrLysAsp..... 1052
||| |||:::|||||::: ||| |||
3205 GTCAAGAGCATCAGAAAGACGAGATACAAATCTGCTCTCTCTTGAG 3254

1053 .AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValC 1069
::: |||:::|||||::: ||| |||
3255 ACTTGAGGCCCCACGTGACTACACGATGAACACCGAGGTTACACCGTCT 3304

1069 ysasp.....ValalaMetAsn 1074
||| |||:::|||||::: ||| |||
3305 GTGCAAGCAAGTCCGCATGTGCCAATGGAAAGTTGAAATTCGCCGCTCC 3354

1075 IleIleMetSerLysSerThrThrTyrSerLeuGluSer..... 1087
||| |||:::|||||::: ||| |||
3355 CCNATCTCGAGAAACCAAGGAATGGACTCTCCGCTCCCACTTCTTGT 3404

```

```

1088 .....ProLysAspProValLeuProAlaArgPhePheThrGlnProA 1102
||| |||:::|||||::: ||| |||
3405 TCTCGCCGACAGATCCCATCATCTCGTCACTCCACGATCAACCAC 3454

1102 sLys.AsnPheSerAsnThrLysAsnTyrLeuPro..... 1113
::: |||:::|||||::: ||| |||
3455 ACCGTGAAGTTTC...AATCTCTCTTACCTTACCTGGGGATCCGAGAG 3501

1114 .....ProGluMetLysSerPhePheThrProGlyLysProL 1126
||| |||:::|||||::: ||| |||
3502 AAGAGCGAAGTCAACGCTCAACTTCAACAATCAAGGAACAGAA 3551

1126 ysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGly 1142
::: |||:::|||||::: ||| |||
3552 GAAATACGAGCGTAACTGGACGCTC.....ACTTCAATGGAA 3589

1143 LysGlnSerGlnThrLysSerArgMetGluThrValSerAsnAlaSe 1159
||| |||:::|||||::: ||| |||
3590 TGCCAGAGTACGAGCTTTTGATCAAGGCCCGCTTAAACAGATCAAC 3639

1159 r.....SerSerSerAsnProSerSerProGlyArgIleLysGlyA 1173
||| |||:::|||||::: ||| |||
3640 GCTGTTGCTGATACAAAGCTACCCGCTGAACCGAACAAGTTCTT... 3684

1173 rgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyr...Thr 1188
::: |||:::|||||::: ||| |||
3685 .....GCCGTTATTGTTGATCTGTTAAGGCCCTTACAACTACTGGACT 3726

1189 MetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAs 1205
||| |||:::|||||::: ||| |||
3727 GTCTCGAGC.....CGCCAGAGAACACGAGAATGATCCGCT 3764

1205 pLeuValArgSerGluLeuGluLysProArgLysLysLysThrProv 1222
::: |||:::|||||::: ||| |||
3765 CGTCGTTCAACTCACAGTTGAG...CCAATGTCGCGACAATACGTCAACA 3811

1222 alThrGluGlnGlu.....GluLysLeuGlyMetAspLeuThrLys 1236
||| |||:::|||||::: ||| |||
3812 TCACCATCAATCCCAATGAGCGGTTGAGCTCAAGAACGTT...CAA 3858

1237 LeuValGlnGluLysProLysGlySerGlnArgSerArgLysArgG1 1253
::: |||:::|||||::: ||| |||
3859 GTGCCACGTGTGTACTTCCATCCATGCTCAACGTTCCGCTCAAG... 3903

1253 yHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLys... 1268
||| |||:::|||||::: ||| |||
3904 .CACCTACTTAATGAGGTTCCGGATCTGTTTCAAGGTGCAAAAGAAC 3952

1269 .....ArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSer 1283
||| |||:::|||||::: ||| |||
3953 AATCCGTACTTTCGATGATGCTCTCAAC..... 3984

1284 ProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyG1 1300
..... 3984

1300 yThrPro.....LysGluG 1305
||| |||:::|||||::: ||| |||
3985 .ACTCCACTCACCACTGTACTCTGATCGCCAAGATTGCTCCGAGG 4033

1305 luProThrMetLysThrSerLysLysGlySerLysLysSerGlyPro 1321
||| |||:::|||||::: ||| |||
4034 AGCCAACTTTCGCTGTTCTCTCCAAGAGACTGAAAGAAATTC... 4077

1322 ProAlaProGluGlu.....GluGluGluG1 1330
||| |||:::|||||::: ||| |||
4078 .....GAAGAGATGATCATCAAGGTTATCCGTGGAGACACAGAAAT 4118

1330 uGluArgGlnSerGlyAsnThrGluGlnLysSerLysGlnHisA 1347
||| |||:::|||||::: ||| |||
4119 TGTGCTCAACTCCAGAACGAGGAGATTCCGCTTAAGGTTACGGTAAGA 4168

1347 rgVal.....SerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360

```



```

1259 AGCGCGTCAACTTCTTAAGTCCATC.....CAAGAGACTCCATTC 1299
308 AlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAs 324
1300 CCATCTCAAAACC..... 1311
324 pIleHisValProIleArgLeuGluCysValLysPheAla...SerHisC 340
1312 .....ATTGCTGAAGCACTCATCAAGTTCCGCCGAGTCTCGCG 1348
340 ysLeuMetAsnHisPro..... 345
1349 TTTCCAAAGAACACAGCGTTGTCGACAATCTGCTTGGCTCGCGCGGA 1398
346 AspLeuAlaLysAspLeuThrGluTyrLeuLysValArg.....Se 359
1399 TCAGTTGTCCGCGGAATGTTGACTACAAGAAATATCCGTCACCTCGTCGG 1448
359 rHisAspProGluGluAlaIleArgHisAspValIleValSerIleValT 376
1449 CGAGGACAACGCGTGAAGTC.....AAGGAGAAGTTCTCTCCGCGTATTC 1492
376 hrAlaAlaLysLysAsp.....IleLeuLeuVal 385
1493 TGCAACAATACAAAGGATGCTGAGACCACCTTACGAGAAGATTTTGGCACTC 1542
386 Asn.....AspHisLeuLeuAsnPheValArgG1 395
1543 AAGAGCATCGGAACGCTGGACTCGACATCTCCGTCAACCCAGCTCAACGA 1592
395 uArgThrLeuAspLysArg.....TrpArgValArgLysGluAlaMetM 410
1593 GATCATTGTTGACAAGCGCCAATTAATCTTCCAGTCCGCAAGAGCCATTG 1642
410 etGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGly 426
1643 ATGCCTCAGACTCTT..... 1659
427 LysAspAlaLysGlnIleAlaTrpIleLysAspLysLeuHisI1 443
1660 AAGGACACCATGCCAGCG.....AAGATCCAAAAGGTACTCTCTCCCAAT 1703
443 eTyr.....TyrGlnAsnSerIleAspAspArgLeuVal. 455
1704 CTACAGACAGACAATACGAGCGAGATC.....CGCATGCTTGCTT 1747
456 .....GluArgIlePheAlaGlnTyr 462
1748 TGTGGAGATGATGCACACCCGCCAGAGAGTCTCTCTCTGTCGAAGTT 1797
463 MetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyrTy 479
1798 GTTTCCTCAATGGAGAAAGAGACCAACCAAGTTCGCTCTTACCCA 1847
479 rLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMet 496
1848 C..... 1848
496 rPysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeu 512
1849 .....CAGATGATCCGTCACTTTGCCAAAGTCCACCACCCCTGT 1887
513 IleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVa 529
1888 TACCAACGT.....GTTGCTATCGTTTGTCTCAAAGGT 1919
529 lMetValIleThrArgAsnLeuProAspPro..... 539
1920 TCITTCTTCCTCGTGTATCAACCAAGAACAAATGATCGCTTCCTCCT 1969
540 ..GlyLysAlaGlnAspPheMetLysLys..... 548

```

```

1970 ACCTCAACTTCACACTCTTCTTCAAAACTCTTCTCTCGAGCTCAATTT 2019
549 ...PheThrClnValLeuGluAspAspGluLysIleArgLysGlnLeuG1 564
2020 GACTTTGCCGCTATCTTCGAGAAGAACTTCTTCTGCTAAAGGATCTTCA 2069
564 uValLeuValSerProThrCysSerLysGlnAlaGluGlyCysVala 581
2070 CGCTTCTCTTGATGCCGTCTTCGGA.....GGAACTGGA 2104
581 rGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPhe 597
2105 ACAAGTACTTTGCCAAATCGGATTTCTCAGCAACACACATGGCAATAC 2154
598 LeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAs 614
2155 GTTCAGATG.....GCTCTTGAAACTC..... 2178
614 pThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspG 631
2179 .....GAAGCATT..... 2187
631 lyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArg 647
2188 .....GAGAAGGAGTCCACAACCTGTTGTCCGCGGACGCCGTATCCAA 2229
648 AlaGlyLeuGluLeuLysValLeuSerPhe.....Th 659
2230 ACCGGAATCAGCTTCTTAAGAACTTGCCCTGAAAATGAACATCCGTGC 2279
659 rHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaC 676
2280 TCCCCAGCAACTACAACGAGAAGGATGCTTCGCCATGGTCTACCTTC 2329
676 ysLeuLys...MetAsp.....AspGluLysValAla 685
2330 GTTACAAGGACATGGATATATGCCATCTCCCGAGTCGATACCCCACTCAT 2379
686 GluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAs 702
2380 GAGAAGTCTATTGAGAAGTACATCAGCAACGGA...AAAGTTGAG..... 2421
702 pPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerL 719
2422 .TTCAGGAAATCCGCGTCTCTT..... 2445
719 ysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAla 735
2446 .....AACCAAGACACAGAGTTCGAGACTCAT.....CATGCC 2478
736 IlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuH1 752
2479 .....GCTTACTTCTATGAGGTATCCG 2501
752 sLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuVal 769
2502 CAATTC.....CCAACCACTCTTGA.....CTTCCACTCATCG 2536
769 hrIleGlyHisIleAlaLeuAlaProAspGlnPheAlaAlaProTrp 785
2537 TCTCTGGAAGATC.....CCAACGGTCTTTTCGGCTGAAGGA 2574
786 LysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspAr 802
2575 CAATTCTCTCTTGACTTGAGAAACCGAACTGCTTTGACCGTTGAGGC 2624
802 gLeuProGlyLysLysThrThr.....LysLeuTrpValp 814
2625 TCGCCCATCCGTTGCTGCCACCACCGCTTTACGAGATGAGAAATGTTACCC 2674
814 ro.....AspGluGluValSerProGluThr 822
2675 CACTCTTCGAGCAAGGAGTCAAGAGCGTTTCAATCCGTACGCGCTACACC 2724

```











[illegible]

```

seq_name: /SIS1/gcgdata/geneseqn/NA2001.DAT:AAF54420
seq_documentation_block:
ID_ID AAF54420 standard; DNA; 2272 BP.
XX AC AAF54420;
XX AC
XX DT 02-APR-2001 (first entry)
XX DT
XX DE DNA encoding protein of the invention #85.
XX DE
XX KW Secreted; transmembrane; gene therapy; ss.
XX KW
XX OS Unidentified.
XX OS
XX PN WO200078961-A1.
XX PN
XX PD 28-DEC-2000.
XX PD
XX PF 18-FEB-2000; 2000WO-US04342.
XX PF
XX PR 23-JUN-1999; 99US-0141037.
XX PR
XX PR 20-JUL-1999; 99US-0144758.
XX PR
XX PR 26-JUL-1999; 99US-0145698.
XX PR
XX PR 01-SEP-1999; 99WO-US20111.
XX PR
XX PR 29-OCT-1999; 99US-0162506.
XX PR
XX PR 30-NOV-1999; 99WO-US28313.
XX PR
XX PR 02-DEC-1999; 99WO-US28551.
XX PR
XX PR 16-DEC-1999; 99WO-US30095.
XX PR
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR
XX PR 06-JAN-2000; 2000WO-US00376.
XX PR
XX PA (GETH ) GENENTECH INC.
XX PA
XX PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N,
XX PI Gao J, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL,
XX PI Pan W, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX PI Watanabe CK, Williams PM, Wood WI;
XX PI
XX DR WPI: 2001-071395/08.

```







[illegible]



2074 CGTCTCGAAGACAACAGAGACTTTGCCAGTTC.....TGGTGGGA 2111

551 nValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVals 568  
:||| ||| :||||: |||:||||: ||| :|||||  
2115 TGTGTCTGAGCTTGAAACATGCTATTAGGAACAAGACAAGTCTT.... 2160

568 erProThrCysSerCysLysGlnAlaGluGlyCysVala-gGluIleThr 584  
||| :||||: ||| :||||: ||| :||||:  
2161 .....TCGTCCCACTGATACTGCCCTGATATTGTCACTGTTTCG 2199

585 LysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetII 601  
||||:||||: |||:||||: ||| :||||:  
2200 CATTTGTTGGCCAAAACATAGAAGCGTGAGAACACCTTCGCAGCTGGA 2249

601 eLysPheLeuGluGluArgIleAlaProValHisIleAspThrGluSerI 618  
||||:||||: ||| :||||: ||| :||||:  
2250 GAAGTACCTC.....GATCGTCTGTGATG 2272

618 leSerAlaLeuLeuLysGlnValAsnLysSerIleAspGlyThrAlaAsp 634  
:||||: ||| :||||: ||| :||||: |||:||||:  
2273 TTCTTGGAAAAGAAATTGCAA...GATGAAGCATCCCGAAGT..... 2313

635 AspGluAspGluGlyValProThrAspClnAlaIleArgAlaGlyLeuGI 651  
||||: ||| :||||: ||| :||||:  
2314 .....GACAACATTCCTCCAAGCGTTGCCGAGATTCTGATTACATCAA 2357

651 uLeuLeuLysValLeuSerPheThrHisPro.....I 662  
: ||||| ||||| ||||| :||||: ||||| :  
2358 CAAGCTCAAGAACTCTCTGCTTCTCGCAAGAAGAACTTCGCGGTGGAG 2407

662 leSerPheHisSerAlaGluThr..... 669  
: ||||| :||||: ||| :||||: ||| :||||:  
2408 TCGAATACTACCAATTCTTACCAGCGCGAGATGCTGATCGCTACTTGT 2457

670 PheGluSerLeuLeuAlaCysLeuLysMetasp.....AspGluLy 683  
:||||: ||| :||||: ||| :||||:  
2458 TATGATACTCTCCGCGTCATGCTCATCCGAAGACGTTGGAAAAACGAGGG 2507

683 sValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleG 700  
:||||: ||||| :||||: ||| :||||: ||| :||||:  
2508 ACTGTTCAACTTCTCTCAAGACGAGCATGTCCATGACGAGCTCC 2557

700 luLuAspPheProHisIleArgSerAlaLeuProValLeuHisHis 716  
:||||: ||||| :||||: ||| :||||:  
2558 AGAACTTTGCCAACACATCAAG.....GTCTTCAGGCC 2592

717 LysSerLysLysGlyProProArgGlnAlaLysTyrrAlaIleHisCysII 733  
||||:||||: ||| :||||: ||| :||||:  
2593 AAGCGCGAGTCTCTTCCA..... 2610

733 eHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGlup 750  
||| :||||: ||| :||||:  
2611 .....CAGGAAGCCAGAGAACAATCCAG 2632

750 roLeuHisLysSerLeuAsp.....ProSerAsnLeuGlu 761  
:||||: ||||| :||||: ||| :||||:  
2633 ATATCAGACAAGCATGGACACAACTCTCAAGCAGAGGCGGAAGCTCGAA 2682

762 HisLeu.....IleThrProLeuValThrIleGlyHisIleAlaLeuLe 776  
:||||: ||| :||||: ||| :||||:  
2683 AACCTTTCTCAACTTCCGAAGCAGAGACTTTATCGACTCTCTCCCGTGA 2732

776 uAlaProAspGlnPheAlaAlaProTrpLysSerTrpValaIleThrPheI 793  
:||||: ||| :||||: ||| :||||:  
2733 CAAGTTGTACTCGGATGCCGATTCACTCGACTCTTGATCGATGAG... 2778

793 leValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThr 809  
||| :||||: ||| :||||: ||| :||||:  
2779 .....AAGGAAAACCTTTGGCAACATTTGCTCCCTGACCGGAT..... 2817

810 LysLeuTrpValProAspGluGluValSerProGluThrMetValLysII 826  
:||||: ||| :||||: ||| :||||:  
2818 .....ATCGAAGAGTGTGAGATCATGAAGCATATGAGTCTTCGCACTCT 2858

826 eGlnAlaIleLysMetValArgTrpLeuLeuGlyMetLysAsnAsh 843  
::: ||| ::| |  
2859 TGAGCAA.....CATATGAAGAATCAAG 2881

843 iSerLys.....SerGlyThrSerThrLeuArgLeuLeuThrIle 857  
::: ||| ::| |  
2882 AAGCTAAGGTACTAAAGTCAATGACTTGGCGTCGAATTGCCTTAATGTG 2931

858 LeuHis.....SerAspGlyAspLeuThrGluGlnGlyLysIleSer.. 871  
||| ||||| ||| |  
2932 GAACATCCAAACCTCCGACGACATCTTCACCCTCAACAACAGCTCAACGC 2981

872 .....LysProAspMetSerArgLeuArgLeuA 881  
||| ||||| ||| |  
2982 TCGCTGGGCTCAACTTCGCGACATGGTGCTCAACAAAGAGAACAGACTCG 3031

881 laLaGlySerAlaIleValLysLeuAlaGlnGluProCystyrHisGlu 897  
::: ||| ::| |  
3032 AGAGAGCTCACCGCTC...GAACACTTCGCGATCGATTGCCAAGAAACA 3078

898 IleIleThrLeuGluGlnTyrglnLeuCys.....AlaLeuAlaI 911  
::: ||||| ::| |  
3079 GTTACTTTGGATCGAGGACAAGACTCGTCTACTCGAGGATTCGAGCGCGTT 3128

911 eAsnAspGluCysTyrglnValArgGlnValPheAlaGlnLysLeuHisL 928  
||| ||||| ||| |  
3129 GACCAACGATCTTTCCGGTGTC.....ATGAACATTCAAA 3163

928 ysGlyLeuSerArgLeuArgLeuProLeuGluTygMetAlaIleCysAla 944  
::: ||||| ::| |  
3164 GAAGATTGTCATCATGGAACGCGATTGGGA.....GCG 3198

945 LeuCysAlalys.....AspProValLysGI 953  
::: ||||| ||| |  
3199 ATTCAAGCCCAAGCTCGACTCGCTTCACAAAGACCGGATGACATCGAAAG 3248

953 uArgArgAlaHisAlaArgGlnCysLeu...ValLysAsnIleasnValArg 969  
||| ||||| ||| |  
3249 AGAACGACCAACAAGCTCAAGCCATTCGTGAAG.....ATATCA 3289

970 ArgGluTyrlLeuLysGlnHisAlaValSerGluLysLeuLeuSer.. 985  
||| ||||| ||| |  
3290 AGAGAATCCCAACAAAGTCTGGATATCTCTTAACAAGAAAGATTCGTGACAC 3339

985 ..... 985

3340 GAAGCTAAACTCGATGAAGCTGGAGATCTCAACGCTCTCTCGGTATTT 3389

986 .....LeuLeuproGluTyrrValValProTyr 994  
||| ||||| ||| |  
3390 GGATCACTTCCAAGCCCTGGCTTACTGCCACCCAGAGACAAGTTGCT... 3436

995 ThrIleHisLeuLeuAlaHisAspProAspTyrrValLysValGlnAspIl 1011  
||| ||||| ||| |  
3437 .....CAGAGGAAGAGCCACAATCTCT...TGGCTGAAGCTGAGCAACTT 3477

1011 e.....GluGlnLeuLysA 1016  
; : : : : ||| : : :  
3478 CTCNATCAACACGCTGCCATCCGTCAAGAGATTGACGGATACGCTGAGGA 3527

1016 spValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032  
::: ||||| ::| |  
3528 TTACAAGAAGATGCGTCAATGGGAGATCGTGTCACTCAAGATCAAAACCG 3577

1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGI 1049  
::: ||||| ::| |  
3578 ATCCCAATACATGCTCTCTTCTCAACGCTCTTGCTG..... 3613

1049 nThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeur 1066  
||| ||||| ||| |  
3614 .....GACTTCAAGAAGGATGGCAAGAGTT 3638







```

1185 GluAspTyrThrMetSerProLeuProGly.....LysLys 1197
      ||| : : : : : : : : : : : : : : : : : :
3697 GGACGACCACTAACCAATCTTTGGGTATCTCTTATCAGGAAC 3746
      ||| : : : : : : : : : : : : : : : : : :
1197 sSerAspLysArgAspSerAspLeuValArgSerGluLeuLysP 1214
      : : : : : : : : : : : : : : : : : : : :
3747 TGCTTCATTGGTGAATAACAGCGCTTC...TTAGAGATACAAGAT 3793
      ||| : : : : : : : : : : : : : : : : : :
1214 roArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGly 1230
      : : : : : : : : : : : : : : : : : : : :
3794 TGCAATTGAGATTAAGCTATTGAGGAGCAATAACCAAAAGCTTGCA 3843
      ||| : : : : : : : : : : : : : : : : : :
1231 MetAspAspLeuThrLysLeuValGlnGluLysProLysGlySerG 1247
      ||| : : : : : : : : : : : : : : : : : :
3844 ATGGAGAATA.....AAACTTCCTCTGCCCG 3872
      ||| : : : : : : : : : : : : : : : : : :
1247 nArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnP 1264
      : : : : : : : : : : : : : : : : : : : :
3873 AGAAGCAGAAAGAGAAATGGAAGCCTCCGGAACAGAACATCAATCT 3922
      ||| : : : : : : : : : : : : : : : : : :
1264 rProGluGluLysArgLeu..... 1270
      : : : : : : : : : : : : : : : : : : : :
3923 ACTCCGAGGAACCGAGATGATCACAAAAGATATCGTACTCGATCAGGT 3972
      ||| : : : : : : : : : : : : : : : : : :
1271 .....LysGluAspIleLeuGluAsnG 1278
      : : : : : : : : : : : : : : : : : : : :
3973 TCTGATTGCTCATCATATGGGATAAGTACGAGAGACATTTTGAAGATAGA 4022
      ||| : : : : : : : : : : : : : : : : : :
1278 uAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProL 1295
      : : : : : : : : : : : : : : : : : : : :
4023 AGATGATCACAGT.....TTAGAGCAAAAGTCACAAAATCCTCCGA 4063
      ||| : : : : : : : : : : : : : : : : : :
1295 ysProLeuGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
      ||| : : : : : : : : : : : : : : : : : :
4064 AA.....GATAAATCCTTGTCTGAAGATCATTGGTTGTGGAC 4101
      ||| : : : : : : : : : : : : : : : : : :
1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluG 1328
      ||| : : : : : : : : : : : : : : : : : :
4102 AAATGAGATCTCCGATGATTCACGGATCCAAACAGATGCAAAACAA 4151
      ||| : : : : : : : : : : : : : : : : : :
1328 uGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysG 1345
      : : : : : : : : : : : : : : : : : : : :
4152 GAGAAAGTTCTCGAGAGACTTAATTACAGATTACAGAAAGCTCTCAATC 4201
      ||| : : : : : : : : : : : : : : : : : :
1345 lnHisArgValSerArgAlaGlnGlnArgAlaGluSerProGlu.Se 1361
      ||| : : : : : : : : : : : : : : : : : :
4202 TTCATGTGCGCGTTGAAGATTGAAAATCAAAAGTGGAGACAGAGGAAA 4251
      ||| : : : : : : : : : : : : : : : : : :
1361 rSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgP 1378
      ||| : : : : : : : : : : : : : : : : : :
4252 GACGAGAAAGAAAGAAACAGGATCG...AGACGATCAAGGACAGAT 4298
      ||| : : : : : : : : : : : : : : : : : :
1378 roSerLys 1380
      : : : : : : : : : : : : : : : : : :
4299 CAACGAAG 4306

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT: AAC90918

seq\_documentation\_block:

ID AAC90918 standard; cDNA; 6981 BP.

AC AAC90918;

XX 19-MAR-2001. (first entry)

DT C. elegans spectrin beta-chain/fodrin cDNA sequence SEQ ID NO:11.

DE Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;

KW protein-protein interaction; identification; ss.

XX Caenorhabditis elegans.

OS

XX

```

PN WO200073328-A2.
XX
PD 07-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-EP05108.
XX
PR 01-JUN-1999; 99GB-0012755.
XX
PA (DEVG-) DEVGEN NV.
XX
PI Van Crieckinge W, Roelens I, Bogaert T, Verwaerde P;
XX WPI; 2001-016508/02.
XX DR P-PSDB; AAB50652.
XX
PT Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
PT identifying unknown human cDNAs which encode proteins that interact
PT with the human unc-5C protein -
XX
PS Example 4; Page 108-110; 246pp; English.
XX
CC The present invention describes 3 variants of human unc-5C cDNAs
CC (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
CC the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
CC protein-protein-interactions between the unc-5 protein and a variety of
CC different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
CC cDNA are useful in methods for identifying compounds which reduce or
CC inhibit the lethal phenotype associated with the expression of the
CC unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC experiments for identifying unknown human cDNAs which encode proteins
CC that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC AAB50646 to AAB50693 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 6981 BP; 2165 A; 1605 C; 1641 G; 1570 T; 0 other;

```

alignment\_scores:

Quality: 186.50 Length: 1680

Ratio: 0.256 Gaps: 78

Percent Similarity: 43.333 Percent Identity: 17.917

alignment\_block:

US-09-512-581-2 x AAC90918 ..

Align seg 1/1 to: AAC90918 from: 1 to: 6981

```

17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValAr 33
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 CCAACTAGTCTGAGTTCTGATTCAGTATGCGACAAAATCTATAA 281
      : : : : : : : : : : : : : : : : : : : :
33 gArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerG 50
      : : : : : : : : : : : : : : : : : : : :
282 TTCAGTTGACTTGAGGTTTCAGAGGATTGAGAGAT..... 315
      ||| : : : : : : : : : : : : : : : : : :
50 luGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
      ||| : : : : : : : : : : : : : : : : : :
316 .....GAACGTGAACCTC.....GTGCAAAAGAAA 339
      ||| : : : : : : : : : : : : : : : : : :
67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCy 83
      ||| ||| : : : : : : : : : : : : : : : : : :
340 ACATTTACAAAATGGGTCAACTCACACTTG...GTTTCAGTCAGCTGCAA 386
      ||| ||| : : : : : : : : : : : : : : : : : :
83 sLeuAlaAspIlePhe.....ArgI 90
      : : : : : : : : : : : : : : : : : : : :
387 AGTACAGACCTCTACATGGATATGCGAGACGGAATAATGCTCTCCGAC 436
      : : : : : : : : : : : : : : : : : : : :
90 letYrAla.....ProGluAlaProTyrThrSerProAspLys 102
      : : : : : : : : : : : : : : : : : : : :
437 TCCTCGCAGTTCTTTCGCGTGAACGCTCTGCCGAAACGACCGCGGAAA 486
      : : : : : : : : : : : : : : : : : : : :

```



[illegible]



[illegible]







```

3651 ..... |||||
      nGluAspGluGlnAsnSerProLysLysGlyLysArgGlyArgProp 1294
      :||| ||| :||| :||| :||| :||| :||| :|||
3668 TGAGTTTGAGCCAGAGGACTTCAGAAAGTTGTAAGAAAGGTTTGCTG 3717
      :||| :||| :||| :||| :||| :||| :||| :|||
1294 roLysProLeuGlyGlyThrPro.....LysGluGluProThrMet 1308
      ||| ||| :||| :||| :||| :||| :||| :|||
3718 ACATCCCGAGAGAAAGACTAGCCCATATCTCGCGAAGAACACCATG 3767
      :||| :||| :||| :||| :||| :||| :||| :|||
1309 LysThrSerLysLys.....GlySerLysLysSerGlyProPr 1322
      :||| :||| :||| :||| :||| :||| :||| :|||
3768 GCACTGGAGAGCGCCGCGCTGGCTGCACACAAAGTTACCCCTATCCC 3817
      :||| :||| :||| :||| :||| :||| :||| :|||
1322 oAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluG 1339
      :||| :||| :||| :||| :||| :||| :||| :|||
3818 ACTGACTGTCCCAACAAATCTTCGAGAGTCTCCCAACCAACAGCTG 3867
      :||| :||| :||| :||| :||| :||| :||| :|||
1339 LysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
      :||| :||| :||| :||| :||| :||| :||| :|||
3868 GTGGCAGCAGATCAAAAGTGAAAGTGTCTCAGCGG..... 3905
      :||| :||| :||| :||| :||| :||| :||| :|||
1356 AlaGluSerProGluSerSerAlaIle.....GluSerThrG1 1368
      :||| :||| :||| :||| :||| :||| :||| :|||
3906 .....AGCCAGTAGATTTCAGGCACCATCTCGAGAACCCACACGAA 3949
      :||| :||| :||| :||| :||| :||| :||| :|||
1368 nSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerPro 1385
      :||| :||| :||| :||| :||| :||| :||| :|||
3950 ATCCGTCCCACTCAATAATCTTCGAGAGAGTCCGACTGACAGCCCA 3999
      :||| :||| :||| :||| :||| :||| :||| :|||
1385 erGln 1386
      :||| :||| :||| :||| :||| :||| :||| :|||
4000 GAGAG 4004

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAc45991

seq\_documentation\_block:

ID AAC45991 standard; DNA; 4689 BP.

XX AAC45991;

XX 18-OCT-2000 (first entry)

DT Arabidopsis thaliana DNA fragment SEQ ID NO: 48513.

DE Arabidopsis thaliana

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EF1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.











```

826 CGCAGCCATAGCGCTGGCAAGAGCCGACAGCAAGCAAGCAAGCAAGCTGA 875
1235 LysLeuValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysA 1252
876 AGAGAGATCCAAACCAATGACATGCTCGGAAACCAAGAGCCGGAGTC 925
1252 rgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLys 1268
926 CTAGCAGGCATAAAGTAAGAGCAAAAGTCGGAGCAGGAGTCCAGGAGG 975
1269 ArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerPro 1285
976 AGAGTGGAGGAGG..... 990
1285 oLysLysGly.....LysArgGlyArgProProLysProLeuGlyGly 1300
991 ..AAGCGAGGAGTGTGACGAGGGCAGG..... 1017
1300 lyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
1018 .....AGCCAGGAGAGAGCCTCCGCCAGAGTCGGAGCGGAGCAGG 1059
1317 LysLysSerGlyProAlaProGluGlu..... 1327
1060 AGCAAGCGGGCAGCAGGCGGAGCAGGAGCGCCGACCAAGCAAGCAAGG 1109
1328 GluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1343
1110 ACAAGAGGAGAGCAGGAGAGAGCAGAGAGAGCGCCAGTCGCAGT 1159
1344 LysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerPro 1360
1160 CGCAGCGCAGCAAGAGTGGAGAGGAGCAGAAAGCGAGGAGCAAGCGGA 1209
1360 uSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGly 1377
1210 CAGCAAGCGRCAGCAGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1259
1377 rgProSerLysThrProSerProSerGlnProLysLys 1389
1260 CCAGTCCAGATCTCCATCCCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1297

```

seq\_name: /SIS1/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ67602

seq\_documentation\_block:

ID AAQ67602 standard; cDNA; 4868 BP.

XX AC AAQ67602;  
 XX DT 20-NOV-1994 (first entry)  
 XX DE Retinoblastoma protein Ap10.  
 XX KW Retinoblastoma-associated polypeptide; RAP; Ap12; Ap10;  
 XX KW transcription factor; EF2; cell cycle; ss.  
 XX OS Homo sapiens.  
 XX PN W09412521-A.  
 XX PD 09-JUN-1994.  
 XX PF 19-NOV-1993; 93WO-US11310.  
 XX PR 20-NOV-1992; 92US-0979156.  
 XX PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX PI Lee W, Shan B;  
 XX DR WPI; 1994-200185/24.

PT Nucleic acid encoding retinoblastoma-associated polypeptide(s) -  
 PT used for producing prods. for use in studying cell cyclng and  
 PS disregulated cell growth.  
 XX Disclosure; Page 38-40; 77pp; English.  
 XX cDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,  
 CC was cloned by direct screening of cDNA expression libraries using  
 CC purified RB protein as probe. Ap12 showed transcription factor E2F  
 CC and RB binding activities. The 5' and 3' cDNA sequences were  
 CC determined for Ap2 (AAQ67395, AAQ67596), Ap8 (AAQ67597, AAQ67598) and  
 CC Ap15 (AAQ67599, AAQ67600), and full-length encoding sequences for Ap4  
 CC (AAQ67601) and Ap10 (AAQ67601), which also encoded RB-associated  
 XX proteins.  
 SQ Sequence 4868 BP; 1755 A; 837 C; 1165 G; 1111 T; 0 other;

alignment\_scores:

Quality: 189.50 Length: 1535  
 Ratio: 0.275 Gaps: 64  
 Percent Similarity: 44.886 Percent Identity: 19.349

alignment\_block:

US-09-512-581-2 x AAQ67602 ..

Align seg 1/1 to: AAQ67602 from: 1 to: 4868

```

19 ValLysGluIleSerAspLysIle.....Se 27
156 ATGAAGAAGAAATTAGACTCAAACTCCATTTACAGGAGGTACAACATAATGAC 205
27 rLysGluGluMetValArgLysMetValLysThrPheMetA 44
206 CAAATTTGAAGCATGCATAGAAATGGAAAAATAGTTGGGAACTTAAGA 255
44 sMetAspGlnAspSerGluGluLysGluLeuTyr..... 56
256 AAGAAGAACTCAGATTTAGTGAATAATTTCTTTGTGATCAC 305
57 .....LeuAsnLeuAlaLe 61
306 CAGGAGTTACTCCAGAGATGAAACTTCTGAGGCTCAATTCGTGATTT 355
61 uHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValArgL 78
356 AGAAATGTCATGCAGAT.....AAATCATCAGCTGAAGATATTGGAG 396
78 euLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
397 ATAAATGTGGCC..... 407
95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleTh 111
408 .....AAGTGAATGCACAGCTGGAGAGAGAGATTTCTTGATGTGGA 448
111 rArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgT 128
449 AAATCAGCTGAGTAGGATCAGATCGGAGAAAGTAGCATTGAGCATGAAG 498
128 yrPheTyrIleLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCys 144
499 CCTCTACTCTGGAGGCTGACTTAGAGGTAGTTCAACACAGAGAGATGT 548
145 PheGluLeuGluAspSerAsn.....GluIlePheThrGlnLeuTyrAr 159
549 TTAGAAAAGACAAATGAAATAGCAGAGAGGTATTGTGCTGCTGAAGA 598
159 gThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMeth 176
599 AGAACTC...TCAGTGGTCACAAGTGAGAGAAACCCAGCTT.....CGTG 639
176 isMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSer 192

```



alignment_scores:	
Quality:	190.50
Ratio:	0.907
Percent Similarity:	52.764
Percent Identity:	24.372
Gaps:	18
Length:	398

alignment\_block:

Align seg 1/1 to: AAC98914 from: 1 to: 2212

1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro.. 1056  
 ||| ::|||::: |||::|  
 245 CGAGTAATTGTTGAGCAT .....GCCGCGGCCACG 276

1057 .....AspAspAlaLysMetAsnGluLysLeuTyrThr 1068

277 GCGAGATGGCACTTACGGTTCTGGACGCAATGAGATATGGTTATTAGAACAAG 326  
.....  
1068 alCysAspValalaMetAsnIlelleMetSerLysSerThr ThrTyr.. 1083  
|||:::|||||  
327 TGGCCGCAGATAAATATGCCCTCTACTCGCACAGATACAGACTTATTG 376  
.....  
1084 .....SerLeuGluSerPro.LysAspProVal 1092

377 TGGAGAAATTTGTCAAGTCGGTGCAGCTGGCAAGACCTAAAGGATTATATG 426

1093 LeuProAlaArgphepheThrGlnProAsp.....LysAs 1104

427 CGTCAGGCAGGAGAAAGTGACTTATGCAGATGCTACAAGGGACGCCAAAA 476

1104 p pheserASnThrlvsASnTyrlaup 1113

477 TGAAGGGTGATTGAATTGTATCTTATCTGATATGAAAGAGGCTTTGG 526

1113 roProGluMetIysSerPhePheThrProGluIysProIysThrThrAsn 1129

527 AA.....AAGTTGGATGGAACTGAA 546

1130 Vallouclv  
Alava]Asn[ysProf.euserSerA] 1141

595

1141 aGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnA 1158

380 .GGT TCCAGACGACGACCCTGGTCCCTAC TCCAGAAAGCCGAGTCA T TCAAGGT 634

[illegible]



[illegible]

1015	LysAspVal.....	1011	.....	1016	.....	1021	.....	1026	.....	1031	.....	1036	.....	1041	.....	1046	.....	1051	.....	1056	.....	1061	.....	1066	.....	1071	.....	1076	.....	1081	.....	1086	.....	1091	.....	1096	.....	1101	.....	1106	.....	1111	.....	1116	.....	1121	.....	1126	.....	1131	.....	1136	.....	1141	.....	1146	.....	1151	.....	1156	.....	1161	.....	1166	.....	1171	.....	1176	.....	1181	.....	1186	.....	1191	.....	1196	.....	1201	.....	1206	.....	1211	.....	1216	.....	1221	.....	1226	.....	1231	.....	1236	.....	1241	.....	1246	.....	1251	.....	1256	.....	1261	.....	1266	.....	1271	.....	1276	.....	1281	.....	1286	.....	1291	.....	1296	.....	1301	.....	1306	.....	1311	.....	1316	.....	1321	.....	1326	.....	1331	.....	1336	.....	1341	.....	1346	.....	1351	.....	1356	.....	1361	.....	1366	.....	1371	.....	1376	.....	1381	.....	1386	.....	1391	.....	1396	.....	1401	.....	1406	.....	1411	.....	1416	.....	1421	.....	1426	.....	1431	.....	1436	.....	1441	.....	1446	.....	1451	.....	1456	.....	1461	.....	1466	.....	1471	.....	1476	.....	1481	.....	1486	.....	1491	.....	1496	.....	1501	.....	1506	.....	1511	.....	1516	.....	1521	.....	1526	.....	1531	.....	1536	.....	1541	.....	1546	.....	1551	.....	1556	.....	1561	.....	1566	.....	1571	.....	1576	.....	1581	.....	1586	.....	1591	.....	1596	.....	1601	.....	1606	.....	1611	.....	1616	.....	1621	.....	1626	.....	1631	.....	1636	.....	1641	.....	1646	.....	1651	.....	1656	.....	1661	.....	1666	.....	1671	.....	1676	.....	1681	.....	1686	.....	1691	.....	1696	.....	1701	.....	1706	.....	1711	.....	1716	.....	1721	.....	1726	.....	1731	.....	1736	.....	1741	.....	1746	.....	1751	.....	1756	.....	1761	.....	1766	.....	1771	.....	1776	.....	1781	.....	1786	.....	1791	.....	1796	.....	1801	.....	1806	.....	1811	.....	1816	.....	1821	.....	1826	.....	1831	.....	1836	.....	1841	.....	1846	.....	1851	.....	1856	.....	1861	.....	1866	.....	1871	.....	1876	.....	1881	.....	1886	.....	1891	.....	1896	.....	1901	.....	1906	.....	1911	.....	1916	.....	1921	.....	1926	.....	1931	.....	1936	.....	1941	.....	1946	.....	1951	.....	1956	.....	1961	.....	1966	.....	1971	.....	1976	.....	1981	.....	1986	.....	1991	.....	1996	.....	2001	.....	2006	.....	2011	.....	2016	.....	2021	.....	2026	.....	2031	.....	2036	.....	2041	.....	2046	.....	2051	.....	2056	.....	2061	.....	2066	.....	2071	.....	2076	.....	2081	.....	2086	.....	2091	.....	2096	.....	2101	.....	2106	.....	2111	.....	2116	.....	2121	.....	2126	.....	2131	.....	2136	.....	2141	.....	2146	.....	2151	.....	2156	.....	2161	.....	2166	.....	2171	.....	2176	.....	2181	.....	2186	.....	2191	.....	2196	.....	2201	.....	2206	.....	2211	.....	2216	.....	2221	.....	2226	.....	2231	.....	2236	.....	2241	.....	2246	.....	2251	.....	2256	.....	2261	.....	2266	.....	2271	.....	2276	.....	2281	.....	2286	.....	2291	.....	2296	.....	2301	.....	2306	.....	2311	.....	2316	.....	2321	.....	2326	.....	2331	.....	2336	.....	2341	.....	2346	.....	2351	.....	2356	.....	2361	.....	2
------	----------------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	---











```

6920  AGCAGAAACCAAGAAATACAGGAGCTGATCGAAAAAATCTCCGCTGTGT 6368
      :::: |::|::|::|::| ::::: :::: |::|::|
922   heAlaGlnLysLeuLysGlyLeuSerArgLeuArgLeuProLeuGlu  938
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
6969  AITTCAGAGCAA...CACACTTGTCTCAGCAGCTCTCTCTAGTGAGCTTCAA 7015
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
939   TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAr  955
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7016  .....AAGGAAACTGA 7026
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
955   gAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluT  972
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7027  AGCACAC...AAACATTGCATG.....TTAAATATAAAGGAATCTC 7064
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
972   yrLeuLys.....GlnHisAla  977
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7065  TGTCCATCCAGCTCTCCAGATCCCTTGGCAGCTTGCAAACTGAGCATGTT 7114
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
978   AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTy  994
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7115  AAGCTAAATACTCAACTCAGACCCCTCTGAACAAATTTAAGTGTGTATA 7164
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
994   rThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspI  1011
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7165  CCGAACTGCTGCAGTCAAGAAGATCATAGCTTGCATCAA.....GATT 7208
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1011  leGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluIle  1027
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7209  ATGAG.....AAGGACCTTCTCTCT..... 7228
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1028  LeuMetAlaLysAsnGlnAsnSerHisAlaPheIleArgLysMetVa  1044
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7229  .....GAGCAAAAGAGGCATGATCAGCTGGGACTCCAAC 7263
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1044  lGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys.  1060
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7264  GCAGTGTGTTGGAGCAG...CACGGCAGAAAATGTCTGCATTTGTCATCTG 7310
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1061  .....MetAsnGluLysLeuTyrThr  1067
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7311  AGGAAGCTCAAGTTCTGCGAAATGAAATCTTTGAATGAGTTACTTTTAAA 7360
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1068  ValCysAspValAlaMetAsnIleMetSerLysSerThrThrTyse  1084
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7361  .....AAAGCAAAATATAATTACAGAGTGTCCAGGATGACTTTTC 7398
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1084  r.....L  1085
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7399  AGAGGTGCAGCTATTCTTAAATCAAGTAGGATCAACACTGCAAGAAGAGC 7448
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1085  euGluSerProLys.....  1089
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7449  TTGAGCAACAAGAAAGCGTTTATGCGAGTGGTGGAGGAATTTGGAGATCTG 7498
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1090  .....AspProValLeuProAlaArgPheThrGlnProAspLysAs  1104
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7499  CAGCTCGATGCTTAAGAACTCAGTCAAGGCATGCAACAGGAAATAGGCG 7548
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1104  nPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe.  1120
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7549  CATTGCTTCTACCATPACAGCTCTTAAACAAAAGGCTAAAGGCATGTGTC 7598
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1121  ...ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys  1136
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7599  AGTCGAAATACAGCTGAGATACCGGTATATCTTGACACCTTTGAGAGG 7648
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1137  ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGl  1153
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
7649  AAATTGCAAGAGCAAGAAGACAAACAAAGAACTTATGCCAGAAATCGA 7698
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|
1153  uThrValSerAsnAlaSerSer.....SerSerAsnProSerSerP  1167
      ::::: :::: |::|::|::|::|::|::|::|::|::|::|::|

```

```

7699 GCACCGGCCCTAGTGTCTAGTGTAAATGGAGGAAGAAATGCTAGACTTT 7748
1167 roGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSer... 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7749 TAGGCATACCTCAAACTGTTCAAGATGAATCCAAGAACTCCAATCAAG 7798
1183 .....GluAsnGlu..... 1185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7799 ATCAAAATGCTAGAAAATGAATGAATGAACTGAACTGGTCAAGATGATGCCATGCA 7848
1186 .....AspThrThrMetSer..... 1190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7849 CAAAGTGAAAAAGTGCATTTTGCACAGCAAACTACTTAAGCAGAAACG 7898
1191 .....SerProLeuProGlyLysLysSerAspLysArgAsp 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7899 CCGAAGCTGAGCTAAACCATGACAGGTCAAACTACTTAAAAAGCAAGAT 7948
1203 AspSerAspLeuValArgSerGluLeuGlu.....LysProArgG1 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7949 AATCTTCAGGCTGCAATGAAAGAAATAGAAACCTACAGAAAATGTTGC 7998
1216 yArgLysLysThrProValThrGluGlnGluLysLeuGlyMetAspA 1233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7999 CAAAGTGCAGTACCATATAAGAGAAATGACAACTT..... 8038
1233 spLeuThrLysLeuVal.....GlnGluGlnLysProLysGlySerGln 1247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8039 ..AAACTAGGTGGTAAGATTGAAATGGAAAAATAAAGTACTCAAAA 8086
1248 ArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8087 GCAACACACCAAGAGATGTCCTACTTAAAGTCTGTTGGAGATAAAGGA 8136
1264 pProGluGluLysArgLeuLysGluAspIle.....LeuGluA 1277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8137 AGAAGGCTTGCGTAGGTATAAGAGGAATTAGGCGAGCACAGGCAGACA 8186
1277 snGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPro 1293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8187 ACGATACACAGATTGTGTTCCAAAAGATTATCAGAAAGCTTCAACTTTC 8236
1294 ProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysTh 1310
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8237 CCTGTGACTTGTGTTGGTGAAGCGGTATGTCAGAGCACAGCAATGCT 8286
1310 rSerLysLysGlySerLysLysSerGlyProProAlaProGluGluG 1327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8287 TGTGCTCAG.....TCGGA 8303
1327 luGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1343
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8304 AGCCGCTTGGAAAGGAGGTGCACAT.....TATAGAAG 8341
1344 LysGlnHisArgValSerArgAlaGlnGlnArgAlaGluSerProG1 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8342 AATATCATCATTTATCAGGCATGTCAAGTCTGAAGATCGAAGAA 8391
1360 uSerSerAlaIleGluSerThrGlnSer.....ThrProGlnLysGlyA 1375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8392 AACAAAGGCAAAATCTGATGCTCATCTTCTCATCTATGGATCATCACAA 8441
1375 rgGlyArgProSerLysThrProSer.....ProSerGlnPro 1387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8442 GAGGCTCGCCTCAAAACTTACAGACATGGCCCTGTACTCCA 8491
1388 LysLysAsn 1390
8492 GAAAGTCT 8500

_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT74666
documentation block:

```

```
seq_name: /SIDS1/gcgcdata/geneseq/geneseq.NA1997.DAT:AAAT74666
seq_documentation_block:
```



625 IasnlySerIleAspGlyThrAlaAspAspGluAspGluGlyValProT 642  
||||| :::: :::: |||| |  
6142 CAATGAATACTTTACTACTCTAAAGAAGGTGGGGTGAA..... 6181

642 hrAspGlnAlaIleArgAlaGlyLeuGluLeuLysValLeuSerPhe 658  
::: :::: :::: ..AAGAGACCTTCTACCWT 6199

659 ThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAl 675  
:::||| :::: ::||| :::: ::||| ::|||  
6200 CAGAGACCTTCAAAGCAGCAGCTCTCTTCCAGATGGAGAGCTCAGAGA 6249

675 acYseuLysMetAspAspGluLysValAlaGluAlaLeuGlnIleP 692  
||||| ||| :::: ::|||  
6250 GTCTTAAGACCAAGATTTCAGATTGGAAGAGCGCGAAGAGGATAA 6299

692 heLysAsnThrGly.....SerLysIle 699  
::: ||||: ||| :::: ::||| ::|||

6300 GTGAAGCTACTAATGAATAAAGAACTCTACTGCTAGATCTCTCTCTA 6349

700 GluGluAspPheProHisIleArgSerAlaLeu...LeuProValLeuHi 715  
||||| :::: :::: |||| |  
6350 GAGGAGGAGATCTTCAGAACTGCTAGCATTTTGAATCAAACCTGTAAAGCGA 6399

715 shIysSerLysLysGlyProArgGlnAlaLysTyrrAlaIleHisC 732  
::: :::: ::||| :::: ::|||  
6400 AAGGAAAACCTTGCCATTCGAAGCAGCAACTGGTTTCAGAAATTGGAGC 6449

732 ysIleHisAlaIlePheSerserLysGluThrGlnPheAlaGlnIlePhe 748  
::: :::: ::||| :::: ::|||  
6450 AGCTATCACTGACATTAAGAGTAGAGACCATGCATTTCGCCAA..... 6493

749 GluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIlePh 765  
::: :::: .....TCTAAACGCCAAAGAGTAGAGCTGTAANTAAATAGC 6531

765 rProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPhea 782  
::: ||||: |||  
6532 CAGTCTCGCTGAAGAAATAAAGATCTTGACAAAAAGAGATGGATGAATTCA 6581

782 laAlaProTrpLysSerTrpValAlaIleThrPheIleValLysAspLeuLeu 798  
:::||| :::: ::|||  
6582 GAGATTCAAGAGGAATCC.....GTGCAAGAACAGTCTTCCCAT 6619

799 MetAsnAspArgLeuProGlyLysLysThr.....ThrLysLeuTrpVa 813  
:::||| ::||| |||||  
6620 CTAACTGAAGAGTTATGTACATATAAGACTGAACCTCAATGTCTCAAGCA 6669

813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIlel 830  
||| ::||| :::: ::|||  
6670 ACAGAAAAGAGACATCAACAACAACTTGCAAGAAAAGTTAAGGAAGTGG 6719

830 ysmetMetValargTrpLeuLeuGlyMetyLysasnAn..... 842  
::: ::||| ||| ::||| ::|||  
6720 ATGAGCTATTGCCACACTTATCATCTCTAAAGGAACAGCTGGACCAAAATA 6769

843 HisserLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHi 859  
::: ::||| ::|||  
6770 CAGATGGAGCTAAGGAATGAANAAGCTCAGA..... 6799

859 sSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMets 876  
::: ::||| ||||: |||  
6800 ....AACTATGAACCTTCGGAAGAGATGGATATCATCGAAAAAGAAATCT 6845

876 erArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGlu 892  
|| |||||  
6846 CAGTCTCGCTGTTA.....ATCGAAGAACGAG 6871

893 ProCysTyrHisGluIleIleThrLeuGluClnTyrGlnLeuCysAlaLe 909  
||| ::||| ::||| ::|||  
6872 CTCAGCAGGAAGAAGATGATGTTTTCGAACAGTATGATGATATCTTGAG.. 6919

909 uAlaIleAsnAspGluCysTyrGlnValArgGln.....Valp 922



XX xenopus sp.  
 OS WO9913061-AL.  
 XX PD 18-MAR-1999.  
 XX PF 10-SEP-1998; 98WO-US19231.  
 XX PR 11-SEP-1997; 97US-0058645.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX PI Cleveland DW, Goldstein LSB, Sakowicz R, Wood KW;  
 XX WPI, 1999-229233/19.  
 DR P-PSDB; AAY01632.  
 XX Centromere-associated protein-E and related nucleic acid  
 PT  
 XX Claim 13; Page 67-73; 77pp; English.  
 PS  
 XX The present sequence encodes CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, ATPase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic; biologicultral and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; vascular malfunction; inflammatory and immune diseases; angiogenesis; hypertension; restenosis; and fungal infections), also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells.  
 XX  
 SQ Sequence 9626 BP; 3540 A; 1692 C; 2104 G; 2290 T; 0 other;

alignment\_scores:  
 Quality: 192.00 Length: 1587  
 Ratio: 0.280 Gaps: 68  
 Percent Similarity: 43.163 Percent Identity: 18.526

alignment\_block:

US-09-512-581-2 x AAX26819 ..

Align seg 1/1 to: AAX26819 from: 1 to: 9626

134 AsnIleAlarPValLys.....SerTyrAsnIle.CysPheGluLeuG 148  
 ||| ||||| ||| :: ||||| |||  
 4153 AATGCTCGTGGTGAAGAGATGAGCTTCAACATCTTGTGAAGGCATTAG 4202  
 148 luAspSerAsnGluIlePheThrGlnLeuTyrArgPheLeuPheSerVal 164  
 ::|||::||| ::|||::||| ::|||::|||  
 4203 TTCAGAACTGGAGCTACTCTCGTCTCATGTAATAATCTGTG..... 4243  
 165 IleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMe 181  
 ::|||::||| ::|||::||| ::|||::|||  
 4244 .....GAGGGAAGAACTTGAATACACAAAAAACTCAATGCCTTGA 4287  
 181 tSerSerIleleCysGluGlyAspThrValSerGlnGluLeuLeuAspT 198  
 ::|||::||| ::|||::||| ::|||::|||  
 4288 AAGAGAGATATG.....GCCAAATCTGAAGAAAGCGAAGTGTGAAGT 4331  
 198 hrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAla 214  
 ::|||::||| ::|||::||| ::|||::|||  
 4332 CCATGTGGGAATCTAAGAGACAAACAATAGCTCAAGAACAAAGCA 4381  
 214 ..... 214

4382 GAGGAATATTCTAGTAAAGAAAAATCAATTAGTTTAAAGAGGTGTTCAG 4431  
 215 .....TyrAspLeuAlaLysAlaLeuLeuL 223  
 4432 TGGTTCACAGACGCTGTAGATGAGATAGAGTCTCTGAAGACACAGCTAA 4481  
 223 ysArgThrAlaGlnAlaIleGluProTyrIleThrPhePhe..... 237  
 ||| ::| ::| ::| ::| ::| ::|  
 4482 AGGCACACAGAAAGGCTGGAATAAAGATAGAGATTATTTTGAACCTT 4531  
 238 .....AsnGlnValLeuMetLeuGlyLysThrSerIleSerAs 250  
 ||| ::||| ::||| ::||| ::||| ::|||  
 4532 GTACAAACTGCAAAACACCAATTTAGTTGAGGAAAAATTTGAAACTCCATT 4581  
 250 pleuSerGluHisValPheAspLeuLle..... 259  
 ::|||::||| ::|||::||| ::|||::|||  
 4582 GCAAGCTGACCATGAGGAAGACAGCATTTATCGCGCTTCTGAAGAAATGG 4631  
 259 ..... 259  
 4632 AGATAAAGTTCTTGGAGAAAAAATTTGAGCGAAATCAGTATTACTGGAA 4681  
 260 .....LeuGluLeuTyrAsn.....IleAs 266  
 |||||::||| ::|||::||| ::|||::|||  
 4682 AGATTGCAAGAGAAAGGCTGGAATCTCTAACAACACTTGAAATCTCTCA 4731  
 266 pSerHisLeuLeuLeuSerValLeu.....ProGlnLeuGluPheLysL 281  
 ::|||::||| ::|||::||| ::|||::|||  
 4732 GAAAGAGATGGAGACGCTCGCTTCTATAAAGATGACCTGCAACAGAACG 4781  
 281 eulysser..... 283  
 |||::|||  
 4782 TAGAAAGCTTGTGAGTGAACACATTATTCTAAAGAGATATTGACACA 4831  
 284 .....AsnAspAsnGluG 288  
 ::|||::||| ::|||::||| ::|||::|||  
 4832 ACCTAAGCATCATTCAGATACTCAAGCTCAGCTGCAGAAAAACACAGCA 4881  
 288 uArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAsps 305  
 ::|||::||| ::|||::||| ::|||::|||  
 4882 AGAGCTACAGTTGGCTAAGAAATCTTGCA.....ATCGCTGCTTCTGACA 4925  
 305 erGluLeuAlaSerGlnAsnLys.....Pro 313  
 ::|||::||| ::|||::||| ::|||::|||  
 4926 ATTGCTCAATACTCAGAAAGAAACCTCTCGAGATTGTGTGCATCCT 4975  
 314 LeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis..... 326  
 ||| ::| ::| ::| ::| ::| ::|  
 4976 CTGGAAGAAAAAGATATTGTTTAACTGAAGAATTGTCATCAAAAAACTAA 5025  
 327 .....ValProI 329  
 5026 TGAACAGGAAAAATTACTACATGAAAGAATGAACCTTGAACAGCTCAGG 5075  
 329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPro 345  
 ::|||::||| ::|||::||| ::|||::|||  
 5076 TTGAGCTAAAGTCGGAG.....GTGGAACAT...CTGATGAGAGAGTATG 5116  
 346 AspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspPr 362  
 ::|||::||| ::|||::||| ::|||::|||  
 5117 ATCGAATCGAAGTCCTCCTTGTAGTCTTGCAGCATGAGAAACATGATAC 5166  
 362 oGluGlu.....AlaIleArgHisAspValIleValSerIleValT 376  
 |||||::||| ::|||::||| ::|||::|||  
 5167 TGAACACAACTTCTTGTCTCTTAACAGCAG.....ATCAAGTAGTTA 5210  
 376 hrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeu..... 389  
 |||::|||::||| ::|||::||| ::|||::|||  
 5211 CTCAAGAAAAAGAGAGCTGCAACAAACCCATGAACACTTAACAGCTGAG 5260  
 390 .....LeuAsnPhe..... 392  
 |||||::|||  
 5261 GTGACCCTCTAAAAAGAGAAATATTGAATTGGTTTGAATTTTAAAAATGA 5310



[illegible]

4202	ATCTCTGTAAGAAACATGCAATGCTTTCACGAATTCGAGAATTCGACAGCC	4251	
1188	ThrMetSerSerProLeuProGly	1200	.....LysLysSerAspLys 1200
4252	ACAACTAACCAATCTTTTGTGGGTCTCTTATCAGAAACTGCTTCATT	4301	
1200	sArgAspSerAspLeuValArgSerGluLeuGluLysProArgGlyA	1217	
4302	GGTGGATACACAGCGATGGCTC	4348	.....TTAGAGATACAAAGATTGCATTTGA 4348
1217	rgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAsp	1233	
4349	GAATTAAAGCTATTGAGGAGCAATAACCAAAAAGCTTCAATGGAAGAA	4398	.....TTAGAGATACAAAGATTGCATTTGA 4398
1234	LeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerAr	1250	
4399	CTA	4427	.....AAACTCTCTGCCCGAAGAAGCAG 4427
1250	glysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTTPProGluG	1267	
4428	AAGGAGAAATGGAAGCCTCCGGAACAGAACATGAATCTACTCCGAGS	4477	.....TTAGAGATACAAAGATTGCATTTGA 4477
1267	luLysArgLeu	1270	
4478	AAACCGAGATGATCACAAAAGATATCGTACTCGATCAGGTTTCTGATTGC	4527	.....TTAGAGATACAAAGATTGCATTTGA 4527
1271	.....LysGluAspLleLeuGluAsnGluAspGluG	1281	
4528	TCATCATATGGGTAAGTACGAGAGACATTTTGAAGATAGAGATGATCA	4577	.....TTAGAGATACAAAGATTGCATTTGA 4577
1281	nAsnSerProLysLysGlyLysArgGlyArgProProLysProLeuG	1298	
4578	CAGT	4611	.....TTAGAGATACAAAGATTGCATTTGA 4611
1298	lyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGly	1314	
4612	....GGTAATCCTTGTCTGGAAGAGTCAATGGTTGTGGACAACTAGAG	4656	.....TTAGAGATACAAAGATTGCATTTGA 4656
1315	SerLysLysLysSerGlyProProAlaProGluGluGluGluGluG	1331	
4657	ATCTCCGATAGATTACCGGATCCAAACAAAGATGCAACAGAGGAAGT	4706	.....TTAGAGATACAAAGATTGCATTTGA 4706
1331	uArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgV	1348	
4707	TCTCGAGAGACTTAATTCAGATTACAGAGCTCTCAATCTTCATGTCG	4756	.....TTAGAGATACAAAGATTGCATTTGA 4756
1348	alSerArgAlaGlnArgAlaGluSerProGluSerSerAlaTrl	1364	
4757	CCGTTCAAGATTGAAATCAAAGTGGACACAGAGGAGAAAGACGAGAAA	4806	.....TTAGAGATACAAAGATTGCATTTGA 4806
1364	eGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys	1380	
4807	GGAAAGAAACAGAGTACG...AGACGATCAAGGGACACATCAACGGA	4852	.....TTAGAGATACAAAGATTGCATTTGA 4852
seq_name: /SIDS1/jcgdata/geneseq/geneseq/NA1999.DAT:AAAX26819			
seq_documentation_block:			
ID	AAAX26819 standard; DNA; 9626 BP.		
XX			
AC	AAAX26819;		
XX			
XX	22-JUN-1999 (first entry)		
DE	Nucleotide sequence of centromere-associated protein-E (CENP-E)		
KW	CENP-E; centromere-associated protein-E; ATPase activity;		
KW	plus end-directed microtubule motor activity; chromosome con-		
KW	microtubule binding activity; chromosome movement; mitosis;		
KW	cell proliferation; tumor; metastasis; vascular malformation;		
KW	inflammatory disease; immune disease; angiogenesis; hyperten-		
KW	sion; fungal infection; selective herbicide; fungicide;		
KW	insecticide; plant growth regulator; activator; cancer cell		



1833 GAAGAAACACCAGCTCCAGTGTAGAACAGGTGGAATTGGTTGGTTTACATC 1882  
 423 erAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLys 439  
 1883 CAGAAAGCTTTGGTCTCTGTCAGAGAGTTACAGAGGAGAACTCAAG 1932  
 440 LeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValG1 456  
 1933 CTAAGGAAATCCGGNAAGAGAAAGATTGAAACAGCGCCCTGATTGA 1982  
 456 uArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrGluA 473  
 1983 GAAG.....TTAGAAATCATGGAAA 2002  
 473 rgMet.....LysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487  
 2003 AGCTTGTTCAGAAAACCTTCTCTCGAGAACTCCATATCAGATTGCAAT 2052  
 488 Ala.....ValLysAlaLeuAsnGluMetTr 496  
 2053 CCTGAGCTGGAACATCAGAGGAAAGCTGAAACATTTGGAGAA..... 2097  
 496 pLysCysGlnAsnLeuLeuArg.....HisGlnValLysA 508  
 2098 .GCTTCTATGTCACCTGCTCAAGAGAAATCAGCGCTTCATTCTGAGAAG 2146  
 508 sPLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAla 524  
 2147 ACATGTTGATTCCCGTCTCGAGAGTCTACAGAAAACAGC...AAGAAA 2193  
 525 IlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLy 541  
 2194 CTCTCTGAGAAACATGGTACTGGAGAAATCTCTTTTAATGCCAATGT 2243  
 541 sAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAsp..... 555  
 2244 AGAACTTGAAGAACTCAAGTCGAATTTGAAAGCGCTAGAGAAATCATGCC 2293  
 556 .....AspGluLys..... 558  
 2294 ACCTGTGAATGATGACAAGACCACCTTTAACTAGTGAAGAGAAAGCCTA 2343  
 559 .....IleArgLysGlnLeuGluValLeuValSerPr 569  
 2344 CTCTCTCATATTGATCAATGAGAAAAGAAATGAAGATCTGGAGAAAGA 2393  
 569 oThrCysSerCysLys.....GlnAlaGluG 578  
 2394 ACATGCAGAAATTAAGTGAAGTTTGGAAATTAGCGACTGAGAGGGAGT 2443  
 578 lCysValArgGluIleThrLysLysLeuGly.....AsnProLys 591  
 2444 CTTCTCTCAGAAAATTT...GAAGAGTTGGGAGTGTCTTTAAATGCCAAG 2490  
 592 GlnProThrAsnProPheLeuLeuMetIleLysPheLeuLeuGluArgI1 608  
 2491 GAC.....TGTGAGTAGCTAGTTTGTGCAATTTTCAGAGAGTCGGAT 2534  
 608 e.....AlaProValHisIleAspThrGluSerIleSerAlaL 621  
 2535 GAATGGCATGGAATCAACGATCCATCCACTCAAGATGAAATCAATGCC 2584  
 621 eutLeLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAsp 637  
 2585 GGGTGAGAGAAATCAAGTGGAGTGCAGTGCATGATGCTCATCAT 2634  
 638 GluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLy 654  
 2635 GAGATTATTGTTCTGCAAAATGTTCCAGACTGGCTT..... 2673  
 654 sValLeuSerPheThrHisProLysPheHisSerAlaGluThrPheG 671  
 2674 .....GAGAAGAGATT 2683

671 luSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAla 687  
 2684 CTTCTTTGATTGCT.....GAGAACTCAGGATTAAGAGGCT 2721  
 688 AlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePr 704  
 2722 TCT...AAACTGTTAGAGAGCTAGTTTCTGAACTAGAAGAGGAA..... 2763  
 704 oHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLysG 721  
 2764 .....AATATTG 2770  
 721 lyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737  
 2771 GGAACAAGTCAGATTGACTCTCGATTAACTGCAATAAAATACTG... 2817  
 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeu..... 751  
 2818 .....AGACAGGGATCTACCAGGTATTGATGAAGCTTGAATAAT 2858  
 752 .....HisLysSerLeuAspProSerAsn.... 759  
 2859 TCCAGGTATAGTTCCGGTGTGATGAGAACTCCGAGACCAGAGAAATATGC 2908  
 760 .....LeuGluHisLeuIleThrProLeuValThrIle 770  
 2909 ATGATATATTGAATAGACTTGAACATGCAAACTATGCTTTTGAGCATT 2958  
 771 .....GlyHisIleAlaLeu.....LeuAlaProAspGlnPh 781  
 2959 CGGGATCAAAACAGCACTCTGCCATAGACAACTCGTCTCTTATTGAATT 3008  
 781 eAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuL 798  
 3009 TCTTCGGCAACTTAATCAGAAGCTGTAGTATTGAAACAGAGAGAGAAA 3058  
 798 euMetAsnAspArgLeuPro..... 804  
 3059 TCCTTGAGGAAGAACTTGAATCTCAGTGTGAGCAGCTCTCGTTTTCGCGG 3108  
 805 .....GlyLysLysThrThrLysLe 811  
 3109 GATGAGACCAGAAACTTATCTTCGTCAACGGAGAAATTGACCACAAAAGT 3158  
 811 uTrpValProAspGluGluValSerProGluThrMetValLysIleGlnA 828  
 3159 T.....AATCAAGGAGTCAACAGAGAAAAGGTTTGTGATGTGGAA. 3198  
 828 lalleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSer 844  
 3199 ..ATTGAGGATTTTCACAGCAGGTGTGAGTTACGAGATGACTAC... 3243  
 845 LysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAs 861  
 3244 .....ACCATTTCTACAGGGAGA 3260  
 861 pGlyAsp.....LeuThrGluGlnGlyLysIleSerLysProAspMetS 876  
 3261 TAATAACAAAACGTTGGATGAGAAAGCATACTTGACAAAG..... 3300  
 876 erArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGlu 892  
 3301 .....TCAACATTGCAAGTTGGAGAGGAG 3324  
 893 ProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLe 909  
 3325 AAATGTAACACTAGAA.....GATGATATCTCTTT 3353  
 909 uAlaIleAsnAspGluCysTyrGln.....ValArgG 920  
 3354 GCTTCTGTCGGAGACTATATATCAAGCAATCTCATCTTCTGTTGGAGG 3403







PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132866.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.



```

2449 .....GAGGTAAAGTCTCTC 2465
1000 aHisProAspTyrValLysValGlnAspLeuLeuLysAspV 1017
: ||||| : : : : :
2466 TTGTGATCCA.....CTGGACCTGCTTCAATA 2494
1017 allLysGluCysLeuTyrPheValLeuGluLeuMetAlaLysAsnGlu 1033
: : : : : : : : : : : : : : :
2495 TCAGAACCAACATTCGCGTATGATCTATGTTGCATAGCCTCAACAAT 2544
1034 AsnAsnSerHisAlaPheIle.....ArgLysMetValGluAs 1046
: ||||| : : : : :
2545 TCTCTCTCAACTGTTCTAGGAGATATAGAGAGGAGATGGTCTGCAA 2594
1046 nIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnG 1063
: : : : : : : : : : : : : : :
2595 CCTGTATAAACTCAAGAC.....AAGC 2617
1063 LuLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLys 1079
: ||||| : : : : :
2618 ACAAGATTAAAGAACCTTCGGAGGAGCTTAAGAAGCTCTCCACCTCTGTT 2667
1080 SerThrThrTyrSerLeuGlu.....SerProLysAs 1090
: ||||| : : : : :
2668 AAAACTCTCTCTCCAGCTCCATGCACCACTCTCTCTCACACCTCAACA 2717
1090 p...ProValLeuProAlaArgPheThrGlnProAspLysAsnPhes 1106
: ||||| : : : : :
2718 CAAGCCGAAGTGAGCGCTAAGCAGCACCTCTCACTCGACCAACCTTA 2767
1106 eAsnThrLysAsnTyrLeuProGluMetLysSerPheThrPro 1122
: ||||| : : : : :
2768 ATAACCTCACTGAAGCTG.....TTTGAGAACATCTCTGCTCTC 2805
1123 GlyLysProLys...ThrThr.....AsnValLeuGlyAlaValAsnLysP 1137
: ||||| : : : : :
2806 GCAGAAGTAAGAACATCTACCAAGAACTATTGGACAGA...AATCGTC 2852
1137 roLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGlu 1153
: ||||| : : : : :
2853 CGAGAACTTCTACGAGAGATACTGAAAGACACGCGACACATCTTATACG 2902
1154 ThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgI 1170
: ||||| : : : : :
2903 AGAGCTTCACTAACTTCGTGGAATCTAAGCGCGATGATCAACTCTCTT 2952
1170 eLysGlyArgLeuAsp...SerSer...GluMetAspHisSerGluAsnGlu 1185
: ||||| : : : : :
2953 AACGATGAATCTAAAGCTAAGAGCTGGAAGAGGACATCAATAAGCTGAA 3002
1186 AspTyrThrMetSerSerProLeuPro..... 1194
: ||||| : : : : :
3003 GAAGACACTGCAACTGAGCTTCGACCTGTACAAACAGTACAAACTGAAC 3052
1195 .GlyLysLysSerAspLysArgAspAspSerAspLeuVal.....A 1208
: ||||| : : : : :
3053 TCGAGAGACTCTTCGACAGAGAGAGAGAGTTCGGCAAGTATAGATGCGAG 3102
1208 rgSerGluLeuGluLysProArgGlyArg.....LysLysThrProVal 1222
: ||||| : : : : :
3103 ATCAAGAACTTCACTGCTGCTCAAGGAGCAGCTTGAAAGCAAACTCAACTC 3152
1223 ThrGluGlnGluGluLysLeuGlyMetAspLeuThrLysLeuValG 1239
: ||||| : : : : :
3153 ACTGAACATCCGAAACACGCTACTCGAGAACCTCTCAGTGTCTTCAACA 3202
1239 nGluGlnLysProLysGlySerGluArgSerArgLysArgGlyHisThrA 1256
: ||||| : : : : :
3203 AGAAGAA.....GGAGCCGAGATCGCCGAGACAGAACACTCTG 3243
1256 laSerGluSerAspGluGlnGlnTrpProGluGluLysArg..... 1269
: ||||| : : : : :
3244 GAGAACACCAAGATCTTCTCAACACATACAAAGGCGCTCGTCAAGTATTA 3293

```

seq\_name: /SIS1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC45028

seq\_documentation\_block:

ID\_AAC45028 standard; DNA; 5414 BP.

XX AAC45028;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 45029.

XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.



[illegible]



ID AAV21451 standard; DNA; 4940 BP.

XX AAV21451;

XX 23-SEP-1998 (first entry)

XX P. falciparum modified gp190 DNA.

XX gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;  
KW monoclonal antibody; passive immunisation; parasite; ss.

XX Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

XX 10..4929

XX CDS

XX /\*tag= a

XX /product= gp190

XX W09814583-A2.

XX 09-APR-1998.

XX 02-OCT-1997; 97WO-EP05441.

XX 02-OCT-1996; 96DE-4040817.

XX (BUJA/) BUJARD H.

XX Bujard H, Pan W, Tolle R;

XX WPI; 1998-240088/21.

XX P-PSDB; AAW54145.

XX Recombinant production of complete gp190/MSP-1 Plasmodium surface  
PT protein - useful in anti-malaria vaccines, also stabilising genes by  
PT reducing their AT content

XX Example 1; Fig 3c; 48pp; German.

XX This sequence encodes a modified plasmodium falciparum gp190/MSP-1  
CC (merozoite surface protein) which has a reduced AT content resulting in  
CC a higher stability of the protein. Such a protein is useful in  
CC vaccines against malaria or for producing monoclonal antibodies (for  
CC passive immunisation). The complete gp190 protein can now be produced  
CC outside the parasite and has, at least over extended regions, the native  
CC pattern of folding. Larger amounts of the protein can be produced  
CC recombinantly than would be possible using the parasites as source.

XX Sequence 4940 BP; 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

alignment\_scores:

Quality: 193.00 Length: 1405

Ratio: 0.317 Gaps: 65

Percent Similarity: 43.345 Percent Identity: 19.644

alignment\_block:

us-09-512-581-2 x AAV21451 ..

Align seg 1/1 to: AAV21451 from: 1 to: 4940

110 ileThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPhe.. 125

;; ;; ||| ::::: ||| ||::: |||

403 GTGAGAACTATCTCTACTACTCAAGGAGCTCAAGTACCCACAGTTGT 452

126 .....AsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVal.. 139

||::: ;; ||| ::::: |||

453 CGACCTCACTAATCATGTGACACTGTGTGATAACATTCATGGCTTCA 502

139 ysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThr 155

|| ..... ::::: ||::: |||

503 AA.....TATCTGATTGACGGTTACGAGAGATCAATGAA 537

156 GlnLeuTyrArgThrLeuPhe..... 162  
||||: |||  
538 CTCCTGTACAAGTTGAATTTCTACTTCGACTTCGTAAGGCCCAAACTGAA 587  
163 .SerValIleAsnAsnGlyHis.....AsnGlnLysValH 174  
::: ||| ::| ||::: |||  
588 TGACGTTTCGCCCAATGACTATTGTCAAAATTCATTCAATTTGAAGATCA 637  
174 isMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThr 190  
::: ||| ::::: |||  
638 GAGCCAAACGAGTTGGACGTTATTGAAGAAGTTGGTCTTCGGA..... 678  
191 ValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHi 207  
::: ||| ::::: |||  
679 ...TATCGCAAGCCTCTGCACAACATCAAGGACAATGTGGAAAGATGGA 725  
207 slyAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysA 224  
: ::::: |||  
726 AGATTATATTAAGAAGAT.....AAGA 748  
224 rgThrAlaGlnAlaIleGluProTyrIleThrPheAsnGlnVal 240  
: ||| ::| |||  
749 AGACCATCGAGACATT.....AACGAGCTG 774  
241 LeuMetLeuGlyLysThrSerIle..... 248  
::: ||| ::::: |||  
775 ATCGAAGAATCCAAAAGACCATAGACAAAAATAAGAATGCAACCAAGGA 824  
249 .....SerAspLeuSerGluHisValPheAspLeuIleLeuGluL 262  
825 GGAAGAAAAGAGAGTTGTACCGCCCGCCAGTACGAC.....CTGTCCA 868  
262 euTyrAsn.....IleAspSerHisLeuLeuSerValLeuPro 275  
::: ||| ::::: |||  
869 TCTATACCAACAGCTTGGAAGCCCATACCTCATCAGCGTACTGGAG 918  
276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnVa 292  
::: ||| ::::: |||  
919 AAGCGCATAGACACCTCTAAGAAGATGAAAT.....AT 953  
292 llyLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlas 309  
954 CAAAGAACTGCTCGCAAGATTAATGAATTCCTCCCGCCAGCCA 1003  
309 exGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIle 325  
: ||| ::::: |||  
1004 ACTCTGGG..... 1011  
326 HisValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMe 342  
1011 ..... 1011  
342 tAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgS 359  
||| ||::: ||| ::::: |||  
1012 .AACACCCCTAACACCTCTCTGACAAAGAACAAAG.....AAGATAGAG 1054  
359 exHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal 375  
: ||| ::| |||  
1055 AGCAGCAGAAAGAG.....ATCAAGAGATCGCC 1083  
376 ThrAlaAlaLysLysAspIleLeuValAsnAspHisLeuLeuAsnPh 392  
::: ||| ::::: |||  
1084 AAACCATTAAGTTCAACATA.....GATTCTCTCTTACT.. 1119  
392 eValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlam 409  
1119 ..... 1119  
409 etMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAla 425  
1119 ..... 1119



```
155870 GGGGCTTACAATTCCTT..... 155886
155879 863 AspleuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuAr 879
155887 GATCTGATTGAGGAAGCAATC..... 155910
155896 879 gLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrH 896
155911 .....GTTTGTGAAGAGCGGTGATAGTTCGAATACC 155944
155945 896 isGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsn 912
155970 GCAGAGGCTATAAATTCACACCTAC..... 155970
155970 913 AspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysG1 929
155971 ...GCAACCTGGTGGATCCCGCAGGCAATTACAGCTCGATTGCCGATCA 156017
155971 929 yLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuC 946
156018 GGCCTGACCATCCGCAATTCGGTA..... 156042
156042 946 ysaLatysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
156043 .....CATATGATTGAACCATC 156060
156060 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGln.....HisAl 977
156061 ACAAAGATGAACCGCATCTCGCGCAACACCTTCAAGAAACCGCGGAAGA 156110
156110 977 aAlaValSerGluLysLeuLeuSerLeu.....LeuProGluTyrValV 992
156111 ACCGATTCGCCNAATTCGCCGAATGATGCAGATGCCCGCAAGACAAAA 156160
156160 992 al.....ProTyrThrIleHisLeuLeu 999
156161 TCCGCAAAATCATGAAATCCGCAAGAGCGGATTCGATGGAACCCCC 156210
156210 1000 AlaHisAspProAspTyrValLysValGlnAspIle..... 1011
156211 ATCGGCGACGACGACGATTCGCATCTGGGCGACTTCATCGAAGATGCCAA 156260
156260 1012 .....GluGlnL 1014
156261 CAATGTTGCCGCCGCGATCGGCAATGTACACCAGCTCGACGAAGTAA 156310
156310 1014 euLysAspValLysGluCysLeuTyrP.....PheValLeuGlu 1026
156311 CCAAGAAATCCTCGAAAGCTGACACCGGTGAGGCAAAAGTCTCGCT 156360
156360 1027 IleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPhe..... 1039
156361 ATCGGTTTCGGCATCGATATGACACCGCACACACCTGGAAGAAGTCGG 156410
156410 1040 .....IleArgLysMetValGluAsnIle.....LysG 1049
156411 CAGACATTTGACCTACGCGCGNAGCCATCCGCAATCGAGCAAAAG 156460
156460 1049 lnThrLysAspAlaGlnGlyPro..... 1056
156461 CACTCCGCAAGCTCGCGCATCCGCAAGAGCGGCGGTTTGAGAAGTTTC 156510
156510 1057 .....AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAs 1070
156511 TTGACAGCGCAAGACAGCAAGCTGTAACACCAAAACCGCAGGTTTCAAA 156560
156560 1070 pValAlaMetAsnIleIleMetSerLysSerThrTyr..... 1083
156561 TACCTGGGTTTTTCTTACACATAAACACGCTTCCACATATCCACACA 156610
156610 1084 .....SerLeuGluSerProLysAspProValLeuPro 1094
```

```
156611 CTCCTATCCCGAGACCTTTCAGAAATTCGCCAAATCCCCCTAAATCCCA 156660
156660 1095 AlaArgPhe.....PheThrGlnProAs 1102
156661 CCAAGACATTTAGGGGATTTTCCATGAGCACCTTCTTTTCAGCAAAACGCA 156710
156710 1102 pLys.....AsnPheSerAsnThrLysAsnTyr..... 1111
156711 CAAGCCATGATTGCCAAACACATCGACCGTTCCTCCACTATTGAAGTTGGA 156760
156760 1112 .....LeuProGluMetLysSerPhePheThrProGlyLys 1124
156761 TCAGTAATTTGTTGGCAACCGATCGAAGACGATACCTGAACCGTCAAGAA 156810
156810 1125 ProLysThrThrAsnValLeuGlyAlaValAsnLys...ProLeuSerSe 1140
156811 CCGCTTACCTTCGAGACACCGCGCGCTCCCGCTATCCCTCTCTCTCC 156860
156860 1140 rAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerA 1157
156861 ATGTTCAAAGCCCTCTCTGCTCGGACAATGGACAGCCTCTCCGATCCGA 156910
156910 1157 snAlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
156911 ACTCGAACACAGCCTCATCCCGCATCGATTTCAACCTGTTTTCGCGTT 156960
156960 1174 Leu..AspSerSerGluMetAspHisSerGlu..... 1183
156961 TTAGCAACTGAGCATCCCGGATTTACAGCACCTTATGCCGCTACCGCAAC 157010
157010 1184 .....AsnGluAspTyrThrMetSerSerProLeu..... 1193
157011 TGGCTGCGCGAAGACAGACACCTGTCGGAACCTGTGGAACGTAACTG 157060
157060 1194 .....ProGlyLysLysSerAspLysArgAspAspSerAspL 1206
157061 CCAACTGACCGAAAAAGGCTTAAAGTAGAGAAAGCA...TCCGCCCGCG 157107
157107 1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
157108 TCGTTGATGCCACCATTTATTCAGACCGCTGGCAGCAACACGCTCAGGCC 157157
157157 1223 ThrGluGlnGluLysLeuGlyMetAspAspLeuThr..LysLeuValG 1239
157158 ATAGAAGTCGATGAGAGAGGACAGTCAGCGGCCAAACACACAGCAGTAA 157207
157207 1239 lnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
157208 GGACAGCGATGCCGTTGGA.....TCAAGAAAAACGGCTCTACA 157248
157248 1256 AlaSerGluSerAspGluGlnGlnTyrProGluLysArgLeuLysG1 1272
157249 AACTCGGTTACAAACACATACCGCTACCGATCGGAAGCTATATCGAG 157298
157298 1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyL 1289
157299 AAATGCG.....ACATTACCC..... 157314
157314 1289 ysArgGlyArgProProLysProLeuGlyGlyThr..... 1301
157315 .....CGCCCAATGCCATGATGCAAAACACCTGTGCGCGTTGT 157353
157353 1302 .....ProLysGluGluProThrMetLysThrSerLysLysG1 1314
157354 TGAAGGTTTACCGGAAGTAGCAGCGCTCTATGCCGACA...AAGGCTAT 157400
157400 1314 ySerLysLysSerGly 1320
157401 GACAGTCGGGAACCGGC 157419
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV21451
seq_documentation_block:
```











CC isolation of *Neisseria meningitidis* DNA sequences, and AAA81322 to







[illegible]

296 uAlaLysMetPheGlyAlaLysAspSerGluLeuAla.....Serg 310  
||::: ||| :::: ||| :||:::|||||  
89114 TAGCGAGCTGGTCGCATCGACCCGACAACTCGCGCAACTGCTAGGAC 89166

310 InAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsn..... 323  
||::: ||| :::: ||| :||:::|||||  
89164 AGGAAGCCGCCAAGCGCGCACGTCAAACAACAAAAAACATAACAACATGCCCTCCG 89213

324 ...AspIleHisValProIleArgLeuGluCysValLysPheAlaSerHi 339  
||::: ||| :::: ||| :||:::|||||  
89214 ATTTCGGTCAAACACCCTCATGCTGACGCTGGTAGACGGCGCAATCCG 89263

339 sCysLeuMetAsnHisProAspLeuAlaLys.....AspLeuThrGluT 354  
: ||::: ||| :::: ||| :||:::|||||  
89264 CAGCCTCTTGATAAATCCGATTGGGCTGCATATATAGACCTGCCCGATT 89313

354 yfLeu.....Lys 356  
|||||  
89314 ATCTCGCGTTGGACGGTGATTTCGCCTGCCTTGGCAATCTTCCCGAATCG 89363

357 ValArgSerHisAsp.....ProGluGluAla..... 365  
||::: ||| :::: ||| :||:::|||||  
89364 ATTAANAACCATGCGCGCTACCCGAAACCGCTCAGGTTTAGAGTATAT 89413

366 .....IleArgHisAspValI 371  
||| ||| |||:  
89414 GCGCGGCTGCCTTACGAAGAACGATAACCCGAATCTTCATTCACGCG 89466

371 leValSer.....IleValThrAlaLalysLysAspIle..... 382  
||| :::: ||| :||:::|||||  
89464 ACCAATCGGAAGAAATGAACAGCAGCAGTGAAGAAGATTGCGAGAATTTC 89513

383 .....LeuLeuValAsnAspHisLeuLeuAsnPheValAr 394  
||::: ||| :::: ||| :||:::|||||  
89514 CAATTCGCATGAAAACATGCTCAATAG.....TTTAAATACAGCCA 89557

394 gGluArgThrLeuAspLysArgTtpArgValArgLysGluAlaMetMetG 411  
: ::||| ||| :::: ||| :||:::|||||  
89558 AATCGAACAATTAAA.....CAAAAAGCTGCAATCCG 89592

411 lyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLys 427  
||||| :::: ||| :||:::|||||  
89593 GCTTAATGAAAGCGAGAAAAAACTTTTGTGCTGCTGCTGACGCGAAAA 89642

428 .....AspAlaAlaLysGlnIleAlaTrrIleLysAspLy 439  
||::: ||| :::: ||| :||:::|||||  
89643 CAAATTCACGCGCGGATTCGCGCATCGTAAACCGTTATGCCCTCTGAA 89692

439 sleuLeuHisIleTyrTyrgInAsnSerIleAspArgLeuLeuValG 456  
| ::||| | ::|||:  
89693 AGCAATTCACCGCGCTGCCACACAGGACAC..... 89723

456 luArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGlu 472  
: ||::: ||| :::: ||| :||:::|||||  
89724 .....TGCAGAACCCCATCCCCCAAGAGCTTTCACAGCGCATCA 89762

473 argMetLysCysLeuTyrTyrLeuTyrAlaThrLeuaspLeuAsnAlaVa 489  
::: ||| ||| :::: ||| :||:::|||||  
89763 GAGTACCTACTCTGCCACGCTTTCAGGTGGGTCCCAACGAAACCGCTCG 89812

489 lLysAlaLeuAsnGlu.....MetTrpLysCysGlnAsnLeuLeuA 503  
||||| : ||| :::: ||| :||:::|||||  
89813 GCATCTTACCAAGAAAGCAGACNAATGCCAGAACCAANTCAGNAG 89862

503 rgHisGln..... 505  
::::|||  
89863 AATATCAAGCAGCACCCGCTCGGTTAAGCATTTGAAGACACGCGCGCG 89912

505 ..... 505  
89913 CTGCGTCAGCTCATCATGGGTAAAGAACGCGGCTACATCACTACTC 89962



```

111 ||| ||| : : : : : ||| |||
7916 CCGGTTACCTTCAGACACCACCGCGCGCTCCCGCTATCCCTCGTGTCC 7965
1140 rAlaGlyLysGlnSerGlnThrLysSerArgMetGluThrValSerA 1157
: : : : : ||| : : : : : ||| : : : : :
7966 ATGTTCAAAAGCGTCTGTCGGAACAATGCACAGCTCTCCGATCCCGA 8015
1157 snAlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
||: : : : : ||: : : : : ||: : : : : ||: : : : :
8016 ACTCGAACACAGCTCATCCCGCATCCATTTCAACCTGTTTTCGGCTT 8065
1174 Leu.. AspSerSerGluMetAspHisSerGlu..... 1183
||| : : : ||| : : : : : ||| : : : : :
8066 TTGACGAACATGACATCCCGATTACAGCACCTTATGCGCTACCGCAAC 8115
1184 ..... AsnGluAspTyrThrMetSerSerProLeu..... 1193
8116 TGCGTGGCGCAAGACACACCTCTGCCGAACCTTGTGGAACATGATTAAC 8165
1194 ..... ProGlyLysLysSerAspLysArgAspSerAspL 1206
||| ||| ||| : : : ||| : : : : :
8166 CCAACTGACCGAAAGGCTTAAGTAGAAGCA...TCGCGCGCGG 8212
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
: : : : : ||| : : : : : ||| : : : : :
8213 TCGTTGATGCCACCATTTACAGCCGCTGGCAGCAACAGCGCTCAGGCC 8262
1223 ThrGluGlnGluLysLeuGlyMetAspAspLeuThr..LysLeuValG 1239
||| : : : : : ||| : : : : : ||| : : : : :
8263 ATAGAAGTCGATGAAGAAGCAAGCAAGCAGCGGCAACACACCGAGTAA 8312
1239 lngGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
: : ||| ||| : : : ||| : : : ||| : : : |||
8313 GGACAGCGATGCCGCTTGA.....TCAGAAAAACGGGCTCTACA 8353
1256 AlaSerGluSerAspGluGlnIntrProGluGluLysArgLeuLysG 1272
||| : : : : : ||| : : : : : ||| : : : : :
8354 AACTCGGTTACAAACAACATACCCGTACCGTACCGTACCGAAGCTATAC 8403
1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGly 1289
: : : : : ||| : : : : : ||| : : : : :
8404 AAATGCG.....ACATTACCC..... 8419
1289 ysArgGlyArgProProLysProLysProLysGlyGlyThr..... 1301
8420 .....CCGCCAATGCCATGATGATGCAACACCTCTCCCGTTGT 8458
1302 .....ProLysGluGluProThrMetLysThrSerLysLysG 1314
||||| : : : : : ||| : : : : : ||| : : : : :
8459 TGAAGGGTTACCGAAGGTACGACCGCTCTATGCCGACA...AAGGCTAT 8505
1314 ySerLysLysLysSerGly 1320
: : : : : ||| : : : : : ||| : : : : :
8506 GACAGTCGGAAAAACCGGC 8524

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21611

seq\_documentation\_block:

ID AAF21611 standard; DNA; 349980 BP.

XX AAF21611;

XX 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.

DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
ds.

XX Neisseria meningitidis.

XX WO20006791-A1.

```

XX 09-NOV-2000.
XX 08-MAR-2000; 2000WO-US05928.
XX 30-APR-1999; 99US-0132068.
XX 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
FA (GENO-) INST GENOMIC RES.
XX Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
DR Neisseria meningitidis B full length genome sequence and open reading
XX frames are used to detect, treat and prevent Neisserial infections -
PT Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
XX used.
XX Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;
SQ

```

alignment\_scores:

Quality: 195.50 Length: 1676  
Ratio: 0.273 Gaps: 81  
Percent Similarity: 42.780 Percent Identity: 18.795

alignment\_block:

US-09-512-581-2 x AAF21611 ..

Align seg 1/1 to: AAF21611 from: 1 to: 349980

```

13 IleThrTyrProGlyValLysGluIleSerAspLysLysLysG 29
: : : : : ||| : : : : : ||| : : : : :
88107 CTGCTGTTCCGGAGCGGTTTCAGTTCCTTCCGACCGCGTGGT.... 88151
29 uGluMetValArgLeuLysMetValValLysThrPheMetAspMetA 46
||||| |||
88152 .....ATGGTCGTGCGCAAGTGCACGGGCAAA 88179
46 spGlnAspSerGlu.....GluGluLysGluLeuTyrLeu 57
: : : : : ||| : : : : : ||| : : : : :
88180 ACGTATATCCGAAGTCCGTCGCAAGCTAAGAAAAACAGCAGACTG 88229
58 AsnLeuAlaLeuHisLeuAlaSerAspPheLeuLys..... 70
: : : : : ||| : : : : : ||| : : : : :

```



[illegible]















/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC45142 + 162.00 186.90 0.0193 1629  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC58907 + 162.00 174.56 0.0941 5468  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC6278 + 161.50 171.97 0.0312 6633  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA45136 + 161.50 168.84 0.1960 9019  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA413178 + 161.00 174.62 0.0934 4812  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF54723 + 161.00 154.27 1.27 35465  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF20348 + 160.50 132.85 0.0090 756  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA34226 + 160.50 182.85 0.0090 756  
/SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT: AAF5759 + 160.50 171.35 0.1421 6236  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF20844 + 160.50 170.12 0.1662 7033  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA34722 + 160.50 170.12 0.1662 7036  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA55274 + 160.50 170.12 0.1663 7036  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA53491 + 160.50 141.66 6.40 114955  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF21437 + 160.50 135.56 14.00 209273  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AAF07285 + 160.00 177.02 0.0886 3360  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA286968 + 160.00 171.11 0.1464 6002  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC48719 + 159.50 179.01 0.0532 2601  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AA051102 + 159.00 178.94 0.0537 2462  
/SIDSL1/gcgdata/geneseq/geneseq/NA2001.DAT: AAF27646 + 158.50 181.41 0.0391 1817  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF57268 + 158.50 179.48 0.0501 2196  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA286888 + 158.00 179.70 0.0487 2021  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA286968 + 158.00 184.58 0.3385 8916  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC89859 + 158.00 164.09 0.3603 9354  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC85440 + 158.00 162.63 0.4345 10795  
/SIDSL1/gcgdata/geneseq/geneseq/NA1995.DAT: AA099876 + 157.50 169.66 0.1764 5093  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AAF71909 + 157.50 189.66 0.1764 5093  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AAF68754 + 157.50 183.86 0.3710 8996  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA27153 + 157.00 185.61 0.0228 1001  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA27151 + 157.00 179.84 0.0478 1763  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA27150 + 157.00 176.93 0.0695 2347  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA244614 + 157.00 165.94 0.2842 6899  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AAT09349 + 157.00 163.46 0.3910 8806  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AAT60713 + 156.50 177.84 0.0617 2017  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF070229 + 156.50 174.00 0.1012 2943  
/SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT: AAQ70102 + 156.50 169.26 0.1899 4766  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AAQ73500 + 156.50 163.26 0.4006 8438  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA644178 + 156.00 164.56 0.3395 6991  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA544778 + 155.50 186.50 0.0204 763  
/SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT: AAQ10212 + 155.50 173.54 0.1072 2721  
/SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT: AAQ10543 + 155.50 173.34 0.1072 2721  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AAQ09923 + 155.50 186.46 0.2858 5452  
/SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT: AAQ70633 + 155.50 160.69 0.5572 9606  
/SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT: AA46467 + 155.00 165.37 0.3060 5711  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AA464543 + 155.00 165.37 0.3060 5711  
/SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT: AAT17485 + 155.00 165.01 0.3202 5912  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AAT17471 + 155.00 165.01 0.3203 5914  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AAT17506 + 155.00 185.01 0.3203 5914  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC47580 + 154.50 176.98 0.0690 1717  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC76827 + 154.50 174.65 0.0930 2158  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC75782 + 154.50 166.02 0.2813 5035  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC64659 + 154.50 165.00 0.3207 5567  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA64659 + 154.50 157.67 0.8214 11435  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA4343872 + 154.50 156.77 0.9221 12493  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA81520 + 154.50 147.81 2.91 30078  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF21608 + 154.50 132.81 71.77 349980  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AAT58840 + 154.50 117.67 138.89 580073  
/SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT: AAQ67601 + 154.00 175.88 0.0795 1800  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA04441 + 154.00 173.45 0.1085 2284  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AAT93091 + 154.00 169.40 0.1911 3523  
/SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT: AAV46460 + 154.00 164.11 0.3593 5711  
/SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT: AAT17500 + 154.00 163.76 0.3760 5914  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF20843 + 154.00 163.24 0.4021 6225  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA34726 + 154.00 163.24 0.4021 6225  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAF21436 + 154.00 145.53 3.89 35384  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA201425 + 154.00 111.11 322.14 1038602  
/SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT: AA24980 + 153.50 161.27 0.5174 7100  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AAT68771 + 153.50 158.89 0.7022 8970  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA64658 + 153.50 155.50 1.09 12515  
/SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT: AA30434 + 153.00 170.31 0.1624 2751  
/SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT: AAT68270 + 153.00 170.18 0.1650 2784  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC58109 + 153.00 168.04 0.2172 3437  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AA51266 + 153.00 188.04 0.2172 3437  
/SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC58904 + 153.00 163.46 0.3907 5386



OM of: US-09-512-581-2 to: N\_Geneseq\_0601.\* out\_format : pfs

Date: Sep 26, 2001 12:37 AM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cyn2.1/USPTO.spool/US09512581/runat\_25092001\_095747.18412/app\_query.fasta\_1.1470  
-DB=N\_Geneseq\_0601 -OFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINARCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELDP=6.000 -DELDP=7.000 -START=1 -MATRIX=blomsum2  
-TRANS=human40.cdi -LIST=1000 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=0 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09512581 -CGN1\_1\_330 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-512-581-2

Query length: 1391

Database: N\_Geneseq\_0601.\*

Database sequences: 730101

Database length: 313950809

Search time (sec): 187.440000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA28051 +	7193.00	8969.67	0.0	5271	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA28052 +	7193.00	8969.51	0.0	5355	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA7423 +	2932.50	3643.34	0.0	3957	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA18338 +	762.50	941.01	1.8e-195	1217	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAV87629 +	698.00	870.72	1.6e-40	439	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA12559 +	609.00	754.10	4.9e-34	738	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA242861 +	513.00	641.59	9.1e-28	351	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC02817 +	501.00	628.35	5.0e-27	295	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC28058 +	349.00	432.25	4.2e-16	530	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA14598 +	225.00	253.62	3.7e-06	5334	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV21511 +	207.00	228.70	9.1e-05	6755	
/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA34578 +	201.50	217.69	0.0004	10136	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA13177 +	197.50	220.21	0.0003	4842	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAC6851 +	197.00	213.51	0.0006	8789	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV09076 +	197.00	213.51	0.0006	8789	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81482 +	195.50	206.43	0.0016	14652	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA21611 +	195.50	174.10	0.0999	34980	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAV21451 +	193.00	214.38	0.0006	4940	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA45028 +	193.00	213.44	0.0006	5414	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA26819 +	192.00	206.33	0.0016	9626	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA74666 +	191.00	204.37	0.0021	10320	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC98914 +	190.50	219.43	0.0003	2212	
/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAC67602 +	189.50	210.15	0.0010	4868	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC45991 +	188.50	209.28	0.0011	4689	
/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAC90918 +	186.50	202.72	0.0025	6981	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA33708 +	186.00	213.54	0.0006	2271	
/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA754420 +	186.00	213.53	0.0006	2271	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA13176 +	186.00	205.80	0.0017	4851	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV18885 +	185.50	194.52	0.0073	13815	
/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAV28296 +	185.00	211.05	0.0009	2565	
/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA58751 +	184.50	204.08	0.0021	4779	
/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA27558 +	183.50	188.61	0.0155	19307	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70106 +	183.00	197.67	0.0049	7458	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC98079 +	181.50	209.52	0.0011	1939	
/SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AAA90338 +	181.50	190.19	0.0127	12923	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV23545 +	181.00	196.41	0.0057	6605	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA74903 +	181.00	196.41	0.0057	6605	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV01731 +	180.00	206.79	0.0015	2108	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAC86236 +	180.00	198.96	0.0041	4548	

/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT91039 +	180.00	192.77	0.0091	8351	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA01274 +	180.00	192.77	0.0091	8351	
/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAA054841 +	179.50	195.00	0.0068	6306	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT78867 +	178.00	193.38	0.0084	6152	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT77783 +	177.50	192.07	0.0100	6579	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT77782 +	177.50	191.70	0.0104	6819	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC65853 +	177.00	190.93	0.0115	6921	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAD00768 +	177.00	187.51	0.0179	9679	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT46187 +	176.00	195.50	0.0038	2638	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT46186 +	176.00	196.27	0.0057	3586	
/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA058610 +	176.00	193.58	0.0082	3624	
/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA058610 +	176.00	193.58	0.0082	4716	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA24044 +	175.50	191.81	0.0103	5276	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA09326 +	174.50	191.79	0.0103	4677	
/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAT37401 +	174.00	188.41	0.0159	6519	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAT705113 +	174.00	190.94	0.0115	4783	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT78868 +	174.00	189.78	0.0134	5361	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA10741 +	173.50	192.73	0.0091	3771	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA09326 +	173.50	186.77	0.0157	6773	
/SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.AAA05301 +	172.50	187.17	0.0187	5760	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC75801 +	172.50	185.88	0.0220	6536	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAT61982 +	171.50	187.48	0.0179	4940	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT61982 +	171.50	181.92	0.0366	8022	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT58840 +	171.00	138.30	9.84	5800	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA25880 +	170.50	181.37	0.0393	7962	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC06611 +	170.00	220.49	0.0003	161	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA075165 +	170.00	179.06	0.0528	9370	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA075165 +	170.00	179.06	0.0528	9370	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC50687 +	169.00	192.74	0.0091	2169	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31959 +	168.50	189.61	0.0137	2744	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76106 +	168.00	195.01	0.0068	1536	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31965 +	168.00	193.70	0.0081	1746	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31407 +	168.00	192.36	0.0096	1992	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA334567 +	168.00	192.36	0.0096	1992	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA334567 +	168.00	190.12	0.0128	2480	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA38817 +	168.00	188.11	0.0166	3023	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA40200 +	168.00	182.63	0.0334	5175	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAV72945 +	168.00	182.35	0.0347	5321	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA14584 +	167.50	186.83	0.0195	3222	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76700 +	167.50	173.71	0.1049	1168	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76822 +	167.00	181.14	0.0405	5298	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC58909 +	167.00	180.80	0.0423	5480	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA20248 +	167.00	150.12	2.16	1113	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA20248 +	167.00	128.71	33.71	9107	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA80911 +	166.50	180.74	0.0426	5181	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA58846 +	166.50	179.00	0.0532	5471	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA68343 +	166.00	190.36	0.0124	6145	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA68343 +	166.00	185.34	0.0236	3104	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA40172 +	166.00	185.33	0.0236	3106	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA68758 +	166.00	174.72	0.0922	8799	
/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA39681 +	165.50	179.25	0.0516	5306	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70095 +	165.50	175.96	0.0786	7326	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA88575 +	165.00	181.50	0.0387	4002	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA60603 +	165.00	181.50	0.0387	4002	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV17568 +	164.50	189.66	0.0136	1689	
/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA75392 +	164.50	172.72	0.1191	8906	
/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA04326 +	164.50	167.93	0.2203	1425	
/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA075181 +	164.50	167.93	0.2203	1425	
/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV20477 +	164.50	179.66	0.0283	1425	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76647 +	164.00	179.66	0.0489	4237	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81493 +	164.00	165.28	0.3092	1738	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA21607 +	164.00	134.70	15.63	3499	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC58905 +	163.50	176.44	0.0740	5471	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC74783 +	163.50	171.87	0.1329	8364	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC51630 +	163.00	188.18	0.0164	1625	
/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.AAA086237 +	163.00	177.80	0.0621	4500	
/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.AAA010378 +	163.00	175.11	0.0877	5858	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA00769 +	163.00	170.47	0.1590	9241	
/SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.AAA03358 +	162.50	185.96	0.0218	1899	
/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.AAA013114 +	162.50	180.97	0.0413	3099	
/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC58911 +	162.50	175.17	0.0870	5477	
/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAA045339 +	162.50	167.26	0.2400	1190	



4647 TTGCATTTCAGGTGGCAGCGGACAGAGTGAAGGAAAGAACCCGGAACCTCCTTC 4698

197 spThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGln 213  
:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
4697 AGACTTTGCTCTCATGTG ..... ACTGAGCTGTTAAAGAC 4734

214 AlaTyArgPLeuAlaLysAlaLeu.....LeuLysArgThrAlaGlnAl 228  
||| :::: ||| |:::|::: |:::|::: |:::|::: |:::|:::  
4735 AAAAATCATCTCCAGGAAAAGCTGCAGAGTTTGCAAAAGACTCACAGGC 4784

228 alieGluProTyrlleThrPhlePheAsnGlnValLeuMetLeuGlyL 245  
:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
4785 ACTGCTTTTGACAAAATCGAGCTGGAAACCACAAATGCCACAACCTCAATA 4834

245 ys.....ThrserIleSerAspLeuSerGluHisVal 255  
|| :::: |:::|::: |:::|::: |:::|::: |:::|:::  
4835 AAGAAGAAACAATGCTTGTCAAGNAATCTGAAGCCCTGCAGGCCAGACTG 4884

256 PheAspLeuIleLeuGluLeuTyraAsnIleAspSerHisLeuLeuLeuSe 272  
::: |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
4885 AGTGAATCAGATTATGAAGAAGCTGAATGCTCCCAAGCCCTTGGAGGCCG 4934

272 ValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluLu. 288  
::: ||| |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
4935 ACTGTTGGAAAAAGTGTAGTTGCGATTCAGAGCTCAGCTCAACACAGAGG 4984

289 .....ArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300  
||| :::: |:::|::: |:::|::: |:::|::: |:::|:::  
4985 AAGTGCATCAGCTGAGACAGAGCATCGAANAATCGAGTTCCGATTCGAG 5034

301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnC 317  
::: |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5035 CGCGATGAAAGAAGCAGCTGCACATCGCAGAGAAC.....TGAAGA 5078

317 ystyrlleUyargPheAsnAspIleHisValProIleArgLeuGluCys 333  
|:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5079 ACAGGAGCGGGAGATG.....ATTCATCTTAAGGATTAAGTTTGAGAACC 5122

334 ValLysPheAlaSerHisCys.....Le 341  
:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5123 TTGAAAGGAATTCAGATGTCAGAGNAACCAAGGAGCTAGTGATCTT 5172

341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrlleLysVala 358  
||| |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5173 GATG.....CCGAAATTCCAAAGCAGAA..GTAGAGACTCTAAAAACAC 5215

358 rgsrHisapProGluAlaIlelreghHisaspvalIlealSerile 374  
::: |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5216 AAATAGAGAGATGCCAGAGACCTGAAAGTTTTTGAATTAGACCTTGTG 5265

375 ValThrAlaAlaLysLysAspIleLeuValAsn..... 386  
::: |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5266 AGTTAAGTCTGAAAGAAAGAAATCTGACAAACAAATACACAGAAAACA 5315

387 .....AspHisLeuLeuAsnPheValArgGluArg 397  
||| |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5316 AGGTCAGTTGTCAGAACTAGACAAGTTACTCTCTCTATTAAAGTCTGT 5365

397 hrLeuAspLysArgTrp.....ArgValArgLysGluAlaMetMet 410  
|:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5366 TAGAAGAAAAGGAGCAGCAGAGATACAGATCAAGAAAGAACTCTAAAACT 5415

411 GlyLeuAlaGlnIleTyrls.....LystyrAlaLeuGlnSerAlaAl 425  
:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5416 GCATGGAGATGCTTCAGANTCAGTTAAGGAGCTAATGAGCGCATGTAGC 5465

425 AcGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuH 442  
|::: |::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5466 AGCCCTGTGTGTGACCAAGAANAATATCAAGGCCACACAGACAGTCTAG 5515

442 isIleTyrrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458  
:::|::: |:::|::: |:::|::: |:::|::: |:::|::: |:::|:::  
5516 ACCCAACATAGACAGACAGATCAGCTGGAATAATAGCATTTGAAAGACTG 5565











4597 ATGAAGAGGAAAAACACAGAGCTTGTAGTCTCATCAAAAGTGAGTGTCTCCA 4646  
 182 rSerIle...IleCysGluGlyAspThrValSerGln...GluLeuLeuA 197  
 4647 TTGCATTACAGTGGCAGAGCAGAGGTGAAGAAAAGACGGAACCTCTTC 4696  
 197 spThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGln 213  
 4697 AGACTTTGCTCTCGANGTG.....AGTGAGCTCTTAAAGAC 4734  
 214 AlaTyrAspLeuAlaLysAlaLeu.....LeuLysArgThrAlaGlnAl 228  
 4735 AAAACTCATCTCCAGGAAAAAGCTCAGAGTTTGAAAAGGACTCACAGG 4784  
 228 aileGluProTyrIleThrPheAsnGlnValLeuMetLeuGlyL 245  
 4785 ACTGCTCTTGACAAAATGTGAGCTGGAAAACCAAAATTGCACAACTGAATA 4834  
 245 ys.....ThrSerIleSerAspLeuSerGluHisVal 255  
 4835 AAGAGAAAGAAATTCCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 4884  
 256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSe 272  
 4885 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAAGCCTTGGAGGCCG 4934  
 272 rValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu. 288  
 4935 ACTGGTGGAAAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGG 4984  
 289 .....ArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300  
 4985 AAGTGCATCAGCTCAGAGAGCGCATCAGAAACTGAGAGTTCGATTGAG 5034  
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnC 317  
 5035 GCCGATGAAAGAGACAGCTGCATCGCAGAGAAC.....TGAAAGA 5078  
 317 ystYrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333  
 5079 ACGGACGGGGAAGT.....ATTCACTTAAAGTAAAGTTGAGAAC 5122  
 334 ValLysPheAlaSerHisCys.....Le 341  
 5123 TTGAAGGGAATTCAGATGTCAGAGAAACACAGGAGCTAGTATTCTT 5172  
 341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysVala 358  
 5173 GATG.....CCGAGAATTCAAAAGCAGAA..GTAGAGACTCTAAAACAC 5215  
 358 rgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIle 374  
 5216 AAATAGAGAGATGCCAGAGCTGAAAGTTTGTGAATTAGACTTGTGTC 5265  
 375 ValThrAlaAlaLysLysAspIleLeuLeuValAsn..... 386  
 5266 ACGTTAAGCTGAAAAGAAAATCTGCACAAAACAAATACAGAAAAACA 5315  
 387 .....AspHisLeuLeuAsnPheValArgGluArgT 397  
 5316 AGGTCACTGTGCAGAACTAGACAAAGTTACTCTCTTCAATTTAAAGTCTGT 5365  
 397 hrLeuAspLysArgTrp.....ArgValArgLysGluAlaMetMet 410  
 5366 TAGAAGAAAAGGACGACGAGATACAGATCAAGAAAGAAATCTAAAAC 5415  
 411 GlyLeuAlaGlnIleTyrLys.....LysTyrAlaLeuGlnSerAlaAl 425  
 5416 GCAGTGAGATGCTTCAGATCACTTAAGGAGCTTAATGAGGCACTAGC 5465  
 425 aGlyLysAspAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuH 442  
 5466 AGCCTTGTGTGGTACCAGAAAATTTATGAGGCCACAGACAGAGTCTAG 5515

442 isileTyrTyrGlnAsnserIleAspAspArgLeuLeuValGluArgIle 458  
 5516 ACCCACCACATAGAGAAAGCATCAGCTGAGAAATAGCATTTGAAAAGCTG 5565  
 459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475  
 5566 AGAGCCCGC.....CTAGAAGCTGATGAAAAGAAAGCA 5597  
 475 sCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaL 492  
 5598 G.....CTCTGTGTCTTCAACCAAC 5617  
 492 euAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAsp 508  
 5618 TGAAGGAAAGTGACATCATCCAGATTACTTAAGGGTAGAGTGAGAAC 5667  
 509 Leu.....LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522  
 5668 CTTGAAAGAGAGCTAGAGTAGCCAGACAAACCAAGAGCATGCAGCTCT 5717  
 522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAsp 539  
 5718 TGAGGCAGAGAAATTCCAAAGGAGAGGTAGAGACCTTAAAGCAAAAATAG 5767  
 539 roGlyLysAlaGlnAspPhe..... 545  
 5768 AAGGATGATCCCAAGTCTGAGAGCTGGAAATTAGATTGTTGTACTATA 5817  
 546 .....MetLysLysPheThrGlnValLeuGluAspAspGluLysI 559  
 5818 AGGTCAGAAAAGAAATCTCACAAATCAATTACAAAAGAGCAAGACG 5867  
 559 eArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnA 576  
 5868 AATATCTGAATTAGAAATAATAAATTCA.....TCAT 5899  
 576 laGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGln 592  
 5900 TTGAAAATATTTTGCAGAAAAGAGCAAGAGAAAGTACAGATGAAGAA 5949  
 593 ProThrAsnProPheLeuGluMet.....IleLysPheLeuLe 605  
 5950 AAATCAAGCACTGCCATGGAGATGCTTCAAAACACAATTTAAAGAGCTCAA 5999  
 605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuI 622  
 6000 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAGCCTGTAAAGGCC.... 6045  
 622 leLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlu 638  
 6046 ..AAGAGCAGAACTTAGT.....AGTCAAGTAGAG 6075  
 639 GlyValProThrAspGlnAla...IleArgAlaGlyLeuGluLeuLeuLys 654  
 6076 TGTCTTGAACCTTGAGAGGCTCAGTGTCTACAGGCTTGTAGAGGCCAA 6125  
 654 sValLeuSerPheThrHisProIleSerPheHisSerAla..... 667  
 6126 A.....AATAATTATATTGTTTTCGCAATCTTCACTGTAAGTGGCC 6163  
 668 .....GluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGlu 682  
 6164 TCATTCAAGAAAGTAGAAGTGGCAAGCAGAAACTGGGAAGAAGAGTAA 6213  
 683 LysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysI 699  
 6214 GAAATCAGTAGACTGAAAAATCAAATT..... 6240  
 599 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal.....L 714  
 6241 ...CAAGACCAAGAGCAGCTTGTCTTAAACTGTCCAGGCTGGAAGGAG 6286







2532 CTCTGGAAGATCCCAACCGCTCATCTCGGCTGAAGGACAAATCTCTCTTG 2581  
 784 rotPLysSerTrp.....Val 789  
 2582 AACTCGAGGAACGAACTTCGTTTGACCGTTGAGGCTCGCCCATCTGTT 2631  
 790 AlaThrPheLeuValLysAspLeuLeuMetAsnAspArgLeu.....Pr 804  
 2632 GCTGCCACCCACGCTCTACGAGATGAGAATGTTCACCCCACTCTTCGAGCA 2681  
 804 oGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProG 821  
 2682 AGGAGTCAAGAGCGTT.....CAATCGGTCGCGCCT 2713  
 821 LuThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837  
 2714 ACACCCCAATCAAGATTCAGCCGTC.....GCT 2742  
 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLe 854  
 2743 GGAATGAAGAGAACTTCGAA..... 2763  
 854 uThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIle 871  
 2764 .....ATCGTCTAC.....AAGGTTG 2779  
 871 erLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVal 887  
 2780 TCGTTCAGAGAACCAAAA.....TCGATCGTC 2808  
 888 LysLeuAlaGlnGluProCysTyr..... 895  
 2809 TCCCTCACCACCTCGCCAGTGTCTCTCTGTTTCCAGGATTCGAA 2858  
 896 HisGluIleLeuThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaLea 912  
 2859 GTTCGAATACATTGAGGCTGAAGAGAACCGTCGTTGTTCCCAATGGC 2908  
 912 snAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928  
 2909 AACAAAGACTCAGGAATTCAGAAAGCTCTC..... 2940  
 929 GlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLe 945  
 2940 ..... 2940  
 945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysL 962  
 2941 .....AACTTCC 2947  
 962 euValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla... 977  
 2948 TCGGACTTGAGTCTCCACCGCGGAACATCCCTCAACCAACACACTCTT 2997  
 978 .....AlaValSerGluLysLeuLe 984  
 2998 GAAAACTGGCTTCTTCGAGCAAGACTTCGAGCTTTCGCTTGAGAACAA 3047  
 984 uSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001  
 3048 GAACCGTCCAGCCGAATTCAGTCCGCTCTCACCCTGGCAACTC.... 3093  
 1001 isAspProAspTyrValLysValGlnAspIleGluGlnLysAspVal 1017  
 3094 .....GAGAAGACCGAACTCTCCCAATC 3117  
 1018 LysGluCysLeuTrpPheValLeuIle...LeuMetAlaLysAsnGln 1033  
 3118 AAGTACAACAAGATCTTCGAAAGAGAGTTCGAGCTCGAACAAGAGACAC 3167  
 1033 uAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnT 1050  
 3168 CGAGAGCGCGGCTGAGTACTTCAACAAGATGGTCAAGAACATCCAGAGG 3217

1050 hrLys.....AspAlaGlnGlyProAspAsp 1058  
 3218 AGCAAGGATACAAATCGTCTATCTCTTGAACCTTGAGGCCCCACGTCAC 3267  
 1059 AlaLysMetAsnGluLysLeuTyrThrValCysAsp..... 1070  
 3268 TACACCATGAACACCGAGCTCACCAACGCTGTGACAAAGCAAGTCCGCAT 3317  
 1071 .....ValAlaMetAsnIleIleMetSerLysSerT 1081  
 3318 GTGCCAATGGGAAGTTGAAATTCGCGCTCCCAATCTCTTGAGAAACCA 3367  
 1081 hrThrTrpSerLeuGluSer.....ProLysAspPro 1091  
 3368 AGGAATGGACTCTCCGCTCCCACTTCTTGTTCGCCCCAGAGATGCCA 3417  
 1092 ValLeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnT 1108  
 3418 TCATCCCTCCGTCAACTCCGCGATCAACCCACCGTGGAGTTC...AACT 3464  
 1108 hrLysAsnTyrLeuPro.....ProGlu 1115  
 3465 CTCTCTTACCTCTACCTGGGATCCCAAGAAAGAGCGAAAGTCCACGTC 3514  
 1116 MetLysSerPheThrProLysProLysThrThrAsnValLeuGln 1132  
 3515 ATGCTCAACTCCCAACATCCAAAGAACAGAGAATAAGAGCGTAACATG 3564  
 1132 ValAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys 1149  
 3565 GACCGTCAAT.....TCAATGGAATGCCAGAGTACGAGCTTTT 3602  
 1149 erSerArgMetGluThrValSerAsnAlaSer.....SerSerSer 1162  
 3603 GATCAAGCGCGCGCTCTTAACCAAGATCAACGCTGTGCTGAGTACAAGC 3652  
 1163 AsnProSerProGlyArgIleLysGlyArgLeuAspSerSerGluMe 1179  
 3653 TCACCCGTGAACCCCAACAGTTCTT.....GCCCGTTATTTT 3689  
 1179 tAspHisSerGluAsnGluAspTyr...ThrMetSerSerProLeuProG 1195  
 3690 CGATCTTGTTAAGACCTACAACACTACTGACTGTCTCGAGC..... 3729  
 1195 LysLysSerAspLysArgAspSerAspLeuValArgSerGluLeu 1211  
 3730 ..CGCCCGAGACACACGAGAATGATCGCGTCTGCCAACTCACAGTT 3777  
 1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGlu.... 1226  
 3778 GAG...CCAATGTCGCCCAATACGTCAACATCACCATGCCAATCCCAAT 3824  
 1227 GluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluLysP 1243  
 3825 GGAGCGCATTCAGCTCAAAAACGTT...CAAGTTCACGCTGTCTACCTTC 3871  
 1243 roLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259  
 3872 CATCCATGTCTCAACGTTCCGCTCAAG.....CACCAACTTACTGAGGCT 3915  
 1260 AspGluGlnGlnTrpProGluGluLys.....ArgLeuLysGluAs 1273  
 3916 TCCGGATCTGTCTCAAGGTCGAAAGAACCAACCAATCCGCACTTTCGATGA 3965  
 1273 pIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysA 1290  
 3966 TGTCTTTTACAAC..... 3978  
 1290 rgGlyArgProProLysProLeuGlyGlyThrPro..... 1302  
 3979 .....ACTCCACTCACCACCTGC 3996



```

231 oTyrIleThrPhePheAsnGlnValLeuMetLeuGlyLysThr... 246
   :: ||||| ::||| ::||| ::|||
1089 TTTCCGACACCTCCACCTCCCACTTAAGGAGATCCACGACTCTT 1138
247 .....SerIleSerAspLeuSerGluHisValPheAsp 257
   ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1139 ACGTCAAGGCTGACAGAGATCCCAATCATGTGAGGACGCTCTT... 1185
258 LeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSerVal 274
   ||| ||| ::||| ::||| ::||| ::|||
1186 GCTATTGCTGGAACCAAGAACACAAATCAACACATCTTGTTCACATG.. 1233
274 uProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeu.. 290
   ::||| ::||| ::||| ::|||
1234 .....GAAACGAGGACATTCCTCC 1252
291 .....GlnValValLysLeuLeuAlaLys..... 298
1253 CACTCGGTCAATTCCTTAAGACCATCCAGAGACACCATTCCTCCCA 1302
299 .....MetPheGlyAlaLysAspSerGluLeuAlaSerG1 310
   ::||| ::||| ::||| ::||| ::|||
1303 AGCATCGCTGAGGCACTCATCAAGTTCCGCGAGCTCGCGTCCCAAGAA 1352
310 nAsnLysProLeuTrpGln..CysTyrLeuGlyArgPheAsnAspIleH 326
   ||| ::||| ::||| ::||| ::|||
1353 CAACCTGGTGTCCGTCACAGCTGCTGGCTCGCGCTGG..... 1391
326 sValProIleArgLeuGluCysValLysPheAlaSerHisCys...LeuM 342
   ::||| ::||| ::||| ::||| ::|||
1392 .....ATCAGTTGTCGCGG.....AATTGTTGACTAC 1419
342 eAsnHisProAsp..LeuAlaLysAsp.....LeuThrG1 353
   ::||| ::||| ::||| ::||| ::|||
1420 AAGAACATCGTCCACTCGTCTGTGAGGACCAAGCGTGAGTCAAGGAGAA 1469
353 uTyrLeuLysValArgSerHisAspProGluGluAla...IleArgHis 369
   ::||| ::||| ::||| ::||| ::|||
1470 GTTCCCTCCGCGTATTCACCAACAAATACAGGATGCTGAGACCACTACG 1519
369 sValIleValSerIleValThrAlaAlaLysLysAspIleLeuVal 385
   ::||| ::||| ::||| ::||| ::|||
1520 AGAAGATTTGGCCCTCAAGACCATCGGAACGCTGGACTC..... 1560
386 AsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArg.. 401
   ||| ::||| ::||| ::||| ::|||
1561 ...GACATCTCCGTCACCAACCAACTCAACGAGATCATTTGTGATAAGCGCCA 1607
402 ....TrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrL 417
   ||||| ::||| ::||| ::|||
1608 ACCACTCCGAGTCCGCAAGGAGCCATTCATGCTCCCTCAGACTTCTT... 1653
417 yLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIle 433
   ||||| ::||| ::||| ::|||
1654 .....AAGGACACCATGCCAGC... 1671
434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
   ||||| ::||| ::||| ::||| ::|||
1672 ...AAGATCCAAAGAGTACTCTCCCAATCTACAAGAACAGACAATACGA 1718
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 467
   ||||| ::||| ::||| ::|||
1719 GCCAGAGATC.....CGCATGTTCTTGTGGAGAAATGATGCATA 1759
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
   ::||| ::|||
1760 CCGCCCGCAGAGAGTCCCTCCTC..... 1782
484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500
   ||||| ::||| ::||| ::|||
1783 .....GTCCAAGTTGTCTCCCAATGGAGAGGAGACCA 1817
500 n.....LeuLeuArgHisGlnValLysAspLeuLeuAspLeuI 513

```

```

1818 CCAACAAGTTGCTGCTCTTACCCACCATGATCCGTCCTACTTTGCCATGT 1867
   ::||| ::||| ::||| ::|||
513 leLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
   ::||| ::||| ::||| ::|||
1868 CCACCAACCCA...TGTTACCAACGTGTTGCTATCGTTTCTCTCAAGGTT 1914
   ::||| ::||| ::||| ::|||
530 MetValIleThrArgAsnLeuProAspPro..... 539
   ::||| ::||| ::||| ::|||
1915 CTTTCTTCTCCTGTTATCAACCAACAAGAACAAATGATCGCTTCTCTCCTA 1964
540 .GlyLysAlaGlnAspPheMetLysLys..... 548
   ::||| ::||| ::||| ::|||
1965 CGCTCAACTTCCACTCTCTCTCAAAACCTCTCTCTGGAGCTCAATTTG 2014
549 ..PheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGlu 564
   ||||| ::||| ::||| ::||| ::|||
2015 ACTTTCGCCCTATCTTCGAGAGAAGTCTTCTTCGCAAGGATCTTCAC 2064
565 ValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValAr 581
   ::||| ::||| ::||| ::|||
2065 GCTTCTCTTGATGCCGCTCTCGGA.....GGAACCTGGA 2099
581 gGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheL 598
   ::||| ::||| ::||| ::|||
2100 CAAGTACTTGGCCAAATCGGATCTCTCAGCAGCACATGGACAAATATG 2149
598 euGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAsp 614
   ::||| ::||| ::||| ::|||
2150 TTCAGATG.....GCTCTTGAGAACTC..... 2172
615 ThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspG1 631
   ||||| ::|||
2173 ...GAAAGCCTC..... 2181
631 yThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArg 648
   ::||| ::||| ::||| ::|||
2182 .....GAGAGGAGTCGACCACTGTGTTCGCGGACGCGGTATTCAAA 2224
648 laGlyLeuGluLeuLysValLeuSer.....PheThr 659
   ::||| ::||| ::||| ::|||
2225 CTGGAATCAAGCTTCTTAAGGAATTCGCCAGAAATGAACATCCGTGCT 2274
660 HisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCy 676
   ::||| ::||| ::||| ::|||
2275 CGCCAGCTACCTACACCGAGAGAGTGCCTCGCATCGTCTACCTTCG 2324
676 sLeuLys...MetAsp.....AspGluLysValAlaG 686
   ||||| ::||| ::||| ::|||
2325 TTACAAGGACATGGACTACGCTCTCTTCAATTCAGACAACTTGTG 2374
686 luAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluAsp 702
   ||| ::||| ::||| ::|||
2375 AGAATCTTATCGAGAAGTTCACCAGCAACGGA...AAGGTTGAG..... 2415
703 PheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLy 719
   ||| ::||| ::|||
2416 TTCAGCGAAATCGCGCTCTTCTC..... 2439
719 sLysGlyProProArgGluAlaLysTyrAlaIleHisCysIleHisAlaI 736
   ::||| ::||| ::||| ::|||
2440 .....AACCAAGAGTCGAGTTCGAGACGCAC.....CATGCC.. 2472
736 lePheSerSerLysGluThrGlnPheAlaGlnIlePheProLeuHis 752
   ||| ::||| ::||| ::|||
2473 .....GCTTACTTCTACGAAGCTATCCGC 2496
753 LysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValTh 769
   ||| ::||| ::||| ::|||
2497 AAATTC.....CCAACAACCTCTTGG.....CTTCCACTCACCAT 2531
769 rIleGlyHisIle.....AlaLeuLeuAlaProAspGlnPheAlaAlap 784
   ||| ||| ::||| ::|||

```







```

764 eThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnP 781
:      :      :      :      :      :      :      :      :
7952 TTTGCAATCTTCAGTGAAGGC..... 7973

781 heAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeu 797
7973 ..... 7973

798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValPr 814
|||||:      :      :      :      :      :      :      :
7974 CTCAITCAAGAGTAGAAGATGCGACGAG.....AAACTGAGAGAA 8017

814 oAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysM 831
|||||:      :      :      :      :      :      :      :
8018 GGATGAAGAAATCACT...AGACTGAAAAATCAAAATTCAGACCAAGAGC 8064

831 eMetValArgTrpLeuLeuGlyMetLysAsnHisSer.....Lys 845
|||||:      :      :      :      :      :      :      :
8065 ACCTTGCTCTTAACGTGCCAGGTGGAAGGAGAGACCAACTTTGGAAG 8114

846 SerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGl 862
:      :      :      :      :      :      :      :
8115 GACAAACTTACAACTGAGAATCTGACGCTGGAATTCGACGACAGAGAT 8164

862 yAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuA 879
:      :      :      :      :      :      :      :
8165 CCAAGTGCTACAAATCAAAATGCCTCTTTGCAGGAC..... 8201

879 rgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnLupProCysTyr 895
8201 ..... 8201

896 HisGluIleIleThrLeuGluGlnTrpGlnLeuCysAlaLeuAlaIleAs 912
|||||:      :      :      :      :      :      :      :
8202 .....ACATTAGAAGTGTGCGAGTCTTACAGNAATCTAGA 8239

912 nAspGluCysTyrGlnVal.....ArgGlnValPheAlaGlnLysL 926
:      :      :      :      :      :      :      :
8240 GAATGAGCTTGAATTCACAAAAATGACAAAAATGTCCTTTGTGTAAGAA 8289

926 euHisLysGlyLeuSerArg.....LeuArgLeuProLeuLupTyr 939
|||||:      :      :      :      :      :      :      :
8290 TAAACAAATGACTGCAAGGAACTGACCTGCAGGAGAAATGCATGAG 8339

940 MetAlaIleCysAlaLeuCysAlaLysAspProVal.....LysGl 953
|||||:      :      :      :      :      :      :      :
8340 ATGGCACAGAAAAACAGCAGAGCTGCAAGAAAGAACTCAGTGGAGAGAAA 8389

953 uArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgA 970
:      :      :      :      :      :      :      :
8390 TAGGCTAGCTGGAGAGTTGCACTTACTGTGGAAGAAATAAAGACGACGA 8439

970 rgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeu 986
:      :      :      :      :      :      :      :
8440 AGATCAATTAAGGAGCTCACACTAGAAAATAGTGAATTAAGAAAGAGC 8489

987 LeuPro.....Glu 989
|||
8490 CTAGATTGTCATGCACAAACACAGGTTGGAAGAGGAGGAAAGTAGAGCA 8539

989 uTyrValValProTyrThrIleHisLeuAlaHisAspProAsp..... 1004
| :      :      :      :      :      :      :      :
8540 GGAATAGTAGTAATATCAGCTACGGCTT.....CATGAAGCTGAAAGA 8583

1005 .....TyrValLysVal 1008
|||||:      :      :      :      :      :      :      :
8584 ACACACAGCTTTGCTTTGGACACAAACACAGTATGAGTAGAATC 8633

1009 GlnAspIle...GluGlnLeuLysAspValLysGluCysLeu..... 1021
||| :      :      :      :      :      :      :      :
8634 CAGACATACCGAGAGAAATGACTTCTTAAGAAAGATGCTTCAGTTTACA 8683

```

```

1022 .TrpPheValLeuGluIleLeuMetAlaLysAsnGlu.....AsnAsnS 1036
:      :      :      :      :      :      :      :
8684 GAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAGAAGAGCTCAATAATT 8733

1036 exHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
|| :      :      :      :      :      :      :      :
8734 CATTTGAAGCTTACTCTCAGATTTTGGAAGAATTGAAGAAACCAAG... 8780

1053 AlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyrThrValcy 1069
|||:      :      :      :      :      :      :      :
8781 .....ATGACAATCTAAAATATGTAATTCAGTTGAAGAAGGAAAA 8821

1069 sAspValAla.....MetAsnIleIleMetSerLysSerThrThrT 1083
:      :      :      :      :      :      :      :
8822 TGAACGTGCCAGGAGGAAATGAAGTTGTTGATC.....AAATCTGTA 8865

1083 yrSerLeuGluSerProLysAspProVal.....Leu 1093
:      :      :      :      :      :      :      :
8866 ACAGCTGGAGAGGAAAGAGAGTACTCCAGAAAGAACTCTCTCAACTT 8915

1094 ProAlaArgPheThrGlnProAspLysAsnPheSerAsnThrLys... 1109
||| :      :      :      :      :      :      :      :
8916 CAAGCTGCACAGAGAGAGCAAGAAACAGCTACTGTTATGGATACCAAGT 8965

1110 .AsnTyrLeuProProGluMetLysSerPheThrProGlyLysProL 1126
:      :      :      :      :      :      :      :
8966 CGATGAATTAACAACGTGAGATCAAGAACTGAAAGAAACTCTTGAAGAA 9015

1126 ysThrThrAsnVal.....LeuGlyAla 1133
|||||:      :      :      :      :      :      :      :
9016 AAACCAAGGAGGAGATGAATACTTGGATAAGTACTGTTCTTCTTATA 9065

1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys.... 1148
:      :      :      :      :      :      :      :
9066 AGCCATGAAGAGTTAGAGAAAGCTAAAGAGATCTTTAGACACACAGTGC 9115

1149 .....SerSerArgMetGluThrValSerAsnAlaSerSerSerS 1162
|||||:      :      :      :      :      :      :      :
9116 CCATCTGTGTTTACACGAATCTAAACAGATTTCCCGAGGCTCTCTTTGC 9165

1162 exAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGlu 1178
:      :      :      :      :      :      :      :
9166 TAGGTCCAGTTGTTCCAGGA..... 9185

1179 MetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuPro.. 1194
|||||:      :      :      :      :      :      :      :
9186 .....CCATCTCCAATCCCTTC 9202

1195 .....GlyLysLysSerAspLysA 1201
:      :      :      :      :      :      :      :
9203 TGTACTGAAAAGAGGTTATCATCTGGCCAAAATAAAGCTTCAGGCAAGA 9252

1201 rgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArg 1217
|||:      :      :      :      :      :      :      :
9253 GCACA.....AGATCCAGTGGAAATATGGGAAATGGTAGA 9287

1218 LysLysThrProValThrGluGlnGluLysLeuGlyMetAspPle 1234
|||||:      :      :      :      :      :      :      :
9288 GGACCAACACCTGCTACC..... 9305

1234 uThrLysLeuValGlnGluLysProLysGlySerGlnArgSerArgL 1251
|||:      :      :      :      :      :      :      :
9306 .....CCAGAGAGCTTTTCTAAAAAAGCAAGAAAGCAGTCA 9342

1251 ysArgGly...HisThrAlaSerGluSerAspGluGlnTrpProGlu 1266
||| :      :      :      :      :      :      :      :
9343 TGAGTGGTATTCCCTGCA..... 9362

1267 GluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSe 1283
|||||:      :      :      :      :      :      :      :
9363 .....GAGACACAGGAAGGTACTGAGTTTGGAGCCAGAGG 9397

1283 rProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyC 1300

```



6402 GAAAAAATGAGGAGAAA. ....ACACAGAGCTTGAGTCTCATCAAAG 6445  
 209 nLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLysLeuArgThrA 226  
 6446 TCGAGTGTCTCCATTCGATTCAGTGGCAGAGCGCAGAGTGAAGAAAAGA 6495  
 226 laGlnAlaIleGluProTyrIleThrThrPheAsnGlnValLeuMet 242  
 6496 CGGAACCTCTCTCAG. .... 6509  
 243 LeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuI 259  
 6510 .....ACTTGTCTCTGATGTGAGTGAGCTGTTAAAGACAAAAC 6550  
 259 e. ....LeuGluLeuTyrAsnIleAspSerHisLeuLeu 271  
 6551 TCATCTCCAGGAACCTGCAGAGTTGGAAAGGACTCACAGGCCTGT 6600  
 271 euSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGlu 287  
 6601 CTTTCGAAAAATGTGAGCTGGAACCAACAAATTCACAACTGAATAAAGAG 6650  
 288 GluArgLeuGlnVal. ....ValLysLeuLeuAlaLysMetPheGlyAl 302  
 6651 AAAGAATTCGTTGTCAAGGAATCTCAAGCCTGCAGGCCACAGACTGAGTGA 6700  
 302 aLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrL 319  
 6701 ATCAGATATGAAGACGTAATGTCTCAAGGCCTTCGAGGCCGCTGAG 6750  
 319 eu. ....GlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333  
 6751 TGGAGAAAGGTGAGTTC. ....GCATTGAGGCTGAGCTCA 6785  
 334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350  
 6786 ACACAGGAGGAAGTGCATCAGCTGAGAAGAGGC. .... 6818  
 350 pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367  
 6819 ....ATCGAGAAACATGAGAGTTCGATTCGATTCGAGCCGATGAAAAGAACGACG 6864  
 367 rGHisAspValIleValSerIleValThrAlaLysLysAspIleLeu 383  
 6865 TGCAC. ....ATCCAGAGAACTGAAGAACCGAGCGGGAGATGAT 6908  
 384 LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400  
 6909 TCATCTAAGGATAAAGTTGAGAACTTGAAGGGAATTCGAGATGTGACA 6958  
 400 sArg. .... 401  
 6959 AGAAACACGAGGAGCTAGTGATTCTTGATGCCGAGAAATTCCAAAGCAGAAG 7008  
 402 .....TriArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle 415  
 7009 TAGAGACTCTAAAACACAAATAGAGAGATGGCCAGAACGCTGAAGATT 7058  
 416 TyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAsp. .AlaAlaLys 431  
 7059 TTTGAATTAGACCTTGTCACGCTTAAGGTCTGAAAAGAAAATCTGACAAA 7108  
 431 sGlnIleAlaThrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnS 448  
 7109 ACAATAACAGAAAAACAGGTTCAGTTGTCAGAACTAGACAAAGTTACTCT 7158  
 448 eIleAspAspArgLeuLeuValGluArgTlePheAlaGlnTyrMetVal 464  
 7159 CTTTCATTTAAAGCTGTTAGAGAAAGGAGCAGACAGAGATACAGATC 7208  
 465 ProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTy 481

7209 AAAGAGAATCTAAACCTGCAGTGGAGATG..... 7238  
481 rAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysC 498  
7238 ..... 7238  
498 ySlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLys 514  
|||||:||||:||||:||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7239 .....CTTCAGAAATCAGTTAAGGAGAGCTTAATAGAGCGAGTAGCA 7277  
515 GlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMetVa 531  
||| ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7278 GCCTGTGTGGTGAC.....CAAGAAATATGA 7306  
531 lIleThrArgAsnLeuProAspPro.....GlyLysAlaG 543  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7307 GCCCACAGACAGACTCTAGACCACCACCAANTAGAGGAAGCATCAGCTGA 7356  
543 InAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIle 559  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7357 GAATAGCATTTGAAAAGCTGAGAGCCGCCCTAGAAGCTGATGAAAAG... 7403  
560 ArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAl 576  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7404 ...AACGAGCTGTGCTTACA.....CAACTGAGGAAG 7438  
576 aGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnP 593  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7439 TGAGCATCATCGAGATTCTACTTAGGGTAGAGTGGAGAACCTTGAAGAG 7488  
593 roThrAsnProPheLeuGluMetLileLysPheLeuGluArgIleAla 609  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7489 AG.....CTAGAGATAGCAGGACAAACCAAGAGCATGCAGCT 7526  
610 ProValHISleAspThrGluSerIleSerAlaLeuLysGlnValas 626  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7527 .....CTTGAGCGACAGAATTCCAAAGGAGAGGTAGACACCTAAA 7567  
626 nLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrA 643  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7568 AGCAAAATAGAAGGATGACC..... 7589  
643 spGlnAlaIleArgAlaGluLeuGluLeuLysValLeuSerPheThr 659  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7590 ..CAAGTCTGAGA..GGTCTGGAATTA..... 7613  
660 HisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCy 676  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7614 GATGTGTTACTATTAAGCTGACAAAA....GAAAACTGTACAAATGA 7657  
676 sLeuLysMetAspAspGluLysValAlaGluAlaAlaLeu.....GlnI 691  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7658 ATTACAAAAGAGCAGACGAATATCTGAATTAGAAATATAAATTCAT 7707  
691 lePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArg 707  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7708 CATTTGAAATATTTTGCAGNAAGACAGCAAGAGAAAGTACAGATGAAA 7757  
708 .....SerAlaLeuProValLeuHisLysSerLysLysGln 721  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7758 GAAATAATCAAGCACTGCCATCGAGATGCTTCAACACAAATTAAGAGCT 7807  
721 ypProArgGlnAlaLysTyrrAlaIleHisCysIleHisAlaIlePheS 738  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7808 CAATGAGAGAGTGGA.....GCCCTCATTAATGACCAAGAACGCTGTA 7851  
738 eSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHis..... 752  
|||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:  
7852 AGCCCCAAGACAGCAATCTTAGTAGTCAAGTAGAGTGTCTTGAACTTCG 7901  
753 .....LysSerLeuAspProSerAsnLeuGluHisLeuIl 764  
7902 AAGGCTCATCTGTACAAGGCTTTCATGAGCCGCAAAAATAATTAATTTGT 7951











1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnPr 1101  
 3580 TTACTACGCCCTGAAGCAGAGAAAATTTTACGAAAGAACTCCAACAAAA 3629  
 1101 oAspLysAsnPheSerAsn...ThrLysAsnTyLeuPro... 1113  
 3630 AGATAAAGCGCTCAAGAGATTCGTTGAANAAGCCGATCTTAACGACACAG 3679  
 1114 ..... 1118  
 3680 CGGCTATCATGAAGTGTTCGGATGTCCTTAAGAGAGAGAGCTCAAA 3729  
 1119 PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValas 1135  
 3730 TACCTGCAAGAAGCTAGAGAAAGGCTGTTCGGGATTTGTTGGCTATGCG 3779  
 1135 nLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgM 1152  
 3780 TAAACCCGATGAAGAAAAAGAAATGCCAAAACCTTTATAGCGATTGTA 3829  
 1152 etGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly 1168  
 3830 TCCAAGAATCCAAATATAAGGACACAAAAACAAATCAATTGAGT 3879  
 1169 ArgIleLysGlyArgLeuAspSerSerGlu...MetAspHisSerGluAs 1184  
 3880 AAAACAGAAAGGTTGCATCAAGCAAGCAGCTGCTGGATTAATTAGATGA 3929  
 1184 n.....GluAspTyThrMetSerSerProLeuProGlyLysLysSerA 1199  
 3930 CCTACTATCAAGAGCCATAGACCAATGTTTATAGAGGGCTTGAGCGATA 3979  
 1199 sPlyArg.....AspAspSerAsp 1205  
 3980 GTGAAGGGCGCTAATCTAGGAATTAACGACACAGCTGATGAAGTCAT 4029  
 1206 LeuValArgSerGluLeuGluLysProArg.....GlyArg 1217  
 4030 CTGATTTATAGCGATCTAAGAAACCGTAAAAACCTTTGATACATGCGCGC 4079  
 1217 gLysLysThrProValThrGluGlnGluLysLeuGlyMetasp..... 1232  
 4080 TAAAGGTTATCCATTGTTACCAATGGATTTCAAAATATGGCGGCGATATG 4129  
 1233 .....AspLeuThrLysLeuValGlnGluGln 1241  
 4130 CCACATTAACCCCACTAATGTTGATCGGACAAAATAGCTAGCGATAAT 4179  
 1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGI 1258  
 4180 CCTATT.....TATGCTTCCATAGA 4199  
 1258 userAsp...GluGlnGlnTTPProGluGluLysArgLeuLysGluAspI 1274  
 4200 GCGTGATATTGCCAAGCAATACGAAACAGAAAAACCATTAAGGATAGA 4249  
 1274 leLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArg 1290  
 4250 ATTTAGAA..... 4257  
 1291 GlyArgProProLysProLeuGlyGlyGlyThrProLysGluGluProTh 1307  
 4258 GCTAAATTAGCTAAGGCTTTAGGTGGCAAT.....AAAAAGATGACGA 4301  
 1307 rMetLysThrSerLysLysLysThrLysLysLysSerGlyProProAlap 1324  
 4302 TAAAGAAAAAGTAAAAAATCCACAGCAGAGAGCTTAAAGCAGAAAAAACA 4351  
 1324 roGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340  
 4352 AGATAGACAAAGATGTCGCAGAACTGCCAAGAATATCATGTGAATCGCT 4401  
 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGI 1357



1470 ..... 1470  
277 LeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVa 293  
1471 .....GAAGAAGAAAGATCAAGTGTT 1493  
293 llyLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerG 310  
1494 AGATTTAATC.....AAAGATGAAACCTA..... 1518  
310 lnaAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326  
1519 ..AAAAAAGCTTA.....CTGAACCAACAAAA 1545  
327 ValProIleArgLeuGluCysValLysPheAla..... 337  
1546 GTTCAAGTCGCCTAGATGTTTGAANAACGCTAAACCGATGAAGAAG 1595  
338 ..SerHisCys.....LeuMetAsnHisPro..... 345  
1596 AAACGAGTCCTAAAACTCATAAATGACCTCGAGATTAGAGAGAAATTC 1645  
346 .....AspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358  
1646 GTAAGGAATTAGAGCTTCAANAAGAGCTTCAAGAGTAT..... 1683  
359 SerHisAspProGluAlaIleArgHisAspValIleValSerIleVa 375  
1684 .....AAGATTGTATCAAAACGCCAA 1706  
375 lThrAlaAlaLysLysAspIleLeuLeu.....ValAsnAspHisLeuL 390  
1707 AACAGAAGCTGAGAAAAACAATGCTTGAAGGCTTGTCTAAAGAAGCTA 1756  
390 euAsnPheValArgGluArgThrLeuAsp.....LysArg 401  
1757 TAGAGAGATTGAACAGCAGCGCTAGATTGTTGAAAAACGCTAAAAACC 1806  
402 TrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysL 418  
1807 GATGAAGAACGAACGAGTGCTTGAANAATATATCCCAAGACTTGCAAAA 1856  
418 sTyrAlaLeuGln.....Sera 424  
1857 AGAATATTAGCTGATATGAGCGTCAAGCGTTTACAAGGATTGCGTATCAA 1906  
424 laAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440  
1907 AAGCTAGAAATGAAAAAGAGAACCAAGATGC.....GAGAAATG 1947  
441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluAr 457  
1948 CTCACGCTGAAGCGGAGGAAAAGTTAGAACACAGGTTCTA..... 1989  
457 gIlePheAlaGlnTyrMetValProHisAsnLeuGluThr...ThrGluA 473  
1990 .....GATTGTTTGAANAACGCTAAACCCGATGNAGAAC 2023  
473 rgMetLysCysLeuTyrTyrLeuTyrAlaThrLeu.....AspLeuAsn 487  
2024 GAAAAAAGTGTTTGAAGAATCTCCCTAAAGACTTACAAAGCGATATCTA 2073  
488 AlaValLysAlaLeuAsn..... 493  
2074 GCCAAGAGAGCGTGAAGACTTTAAAGACTCGGTATCTCAAGCCAAAAAC 2123  
494 .....GluMetTrpLysCysGlnAsnLeu..... 502  
2124 CGAAGCTGAGAAAAAAGATGCGAGAAATTTACTCACCCCTGAAGCGGAAA 2173  
503 .....ArgHisGlnValLysAspLeuLeuAspLeu 512  
.....GTGCTAGATTGTTGAAAAACGCTAAA 2883



```

|||||
113 AGCATCCGTGATAAGGATGTCCTTACTGTAGCTGCTGCTGCTGCTGAT 64
87 IlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLysLeu 103
|||||
63 ATTTTCAGGATTATGCTCTGCTGAGCTCTTACACATCCCTGATAAAGT 14
103 uLys 104
|||||
13 AAAG 10

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseq/NAL1998.DAT:AAx14598

seq\_documentation\_block:

ID AAX14598 standard; DNA; 5334 BP.

```

XX AC AAX14598;
XX DT 31-MAR-1999 (first entry)
XX DE H. pylori GHPO 1755 gene.
XX KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX KW peptic ulcer disease; ss.
XX OS Helicobacter pylori.
XX FH Key
XX FT Location/Qualifiers
XX FT CDS 22..5253
XX FT /*tag= a
XX PN M09843478-A1.
XX PD 08-OCT-1998.
XX PF 01-APR-1998; 98WO-US06371.
XX PR 29-JUL-1997; 97US-0902615.
XX PR 01-APR-1997; 97US-0833457.
XX PR 24-JUN-1997; 97US-0881227.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
XX DR P-PSDB; AAW98879.
XX PT New isolated Helicobacter polynucleotides - used to develop products
XX PT for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX PS Claim 1; Page 2000-2008; 2054pp; English.
XX SS
XX CC This sequence represents a polynucleotide of the invention. It was
XX CC isolated from Helicobacter pylori and encodes a H. pylori GHPO protein.
XX CC The polypeptides can be used for preventing or treating Helicobacter
XX CC infections, and gastroduodenal diseases associated with these
XX CC infections, including acute, chronic, and atrophic gastritis, and peptic
XX CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX CC for the production of antibodies. The products can also be used for
XX CC detection and diagnosis.
XX SQ Sequence 5334 BP; 2206 A; 898 C; 1070 G; 1159 T; 1 other;

```

```

alignment_scores:
  Quality: 225.00      Length: 1545
  Ratio: 0.342        Gaps: 71
Percent Similarity: 42.524 Percent Identity: 18.447

```

alignment\_block:

```

US-09-512-581-2 x AAX14598 ..
Align seg 1/1 to: AAX14598 from: 1 to: 5334
4 SerLysThrArgThrAsn...AspGlyLysIleThrTyrProGlyVa 19
|||||
955 ACAGAACTCAACCAATTTTGTGGAGACAAGTCA..... 990
19 llysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuL 36
|||||
991 .GAAGAAATACTGACGACTCTAACGATCAAGAGATTATCAAGGAAGCA 1039
36 ys..... 36
1040 AAAGAAATATATTATTGCTGGCATTGTAGTCCTGCTTCTTATCGTATT 1089
37 .....MetValLysThrPheMetAspMetAspGlnAs 48
1090 ATTTTATTTCTAGAGCAATTTTCTACTACTCTTCATCCCTTGAAGATAA 1139
48 pSer.....GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuH 62
|||||
1140 AAGCTCTCGTTTTCAGAAAGACAGGAATCTTTATGTCATGAAATCC 1189
62 isLeuAlaSerAspPhe.....PheLeuLysHisProGlyLysAspVal 76
|||||
1190 AAATAAGCAAGAGTATACCGATTGCTGAAAGACGGAATGAAAAAGGC 1239
77 ArgLeuLeuValAlaLacCysCysLeuAlaAspIlePheArgIleTyrAla 93
|||||
1240 AATATGATCGAT..... 1251
93 oGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheI 110
|||||
1252 .....AAGATCTTTCTTC..... 1266
110 leThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
|||||
1267 .....AATGACGATCCCAATAGAACC 1287
127 ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
|||||
1288 TTATCACTATTG.....AATATTGCA..... 1311
143 eCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgT 160
|||||
1312 .....GAAATTGAGCAGAAAAACCCGTTG.....AGAG 1339
160 hrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHis 176
|||||
1340 CCTTTATGAATGTATTAGTAATGCTGCACTATGAAGAATGTTTGAAG 1389
177 MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGl 193
|||||
1390 CTTATC..... 1395
193 nGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnL 210
|||||
1396 .....AAGACAAAAAAC 1408
210 euAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLysArgThrAla 226
|||||
1409 TTCAAGATCAG.....ATGAAAAGACTCTTA 1434
227 GlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMetLe 243
|||||
1435 GAGGCT.....TATAACGACTGCATCAAAAA 1460
243 uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleL 260
|||||
1461 TGCCAAACT..... 1470
260 euGluLeuTyrAsnIleAspSerHisLeuLeuSerValLeuProGln 276

```



PN EP1033401-A2.  
 XX 06-SEP-2000.  
 XX 21-FEB-2000; 2000EP-0200610.  
 XX 26-FEB-1999; 99US-0122487.  
 XX (GEST ) GENSET.  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI; 2000-500381/45.  
 XX P-PSDB; AAG02811.  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

alignment\_scores:  
 Quality: 501.00 Length: 97  
 Ratio: 5.165 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-512-581-2 x AAC02817 ..  
 Align seg 1/1 to: AAC02817 from: 1 to: 295  
 833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSe 849  
 1 GTTCGATGGCTACTTGGAAATGAAATAATACAGTAATCAGGAACCTC 50  
 849 rThrLeuArgLeuLeuThrThrLeuHisSerAspGlyAspLeuThrG 866  
 51 TACCTTAAGATTGCTAACACAAATATTGCATAGTGAGACTTGACAG 100  
 866 luGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882  
 101 AACAGGGGAAAATTAGTAAACCCAGATATGTCAGCTGTGAGACTTGC 150  
 883 GlySerAlaIleValLysLeuAlaGlnGluProCystThrHisGlu 899  
 151 GGGAGTGCTATTGTGAAGCTGGCACAAGAACCTGTATACCATGAAT 200  
 899 eThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGly 916  
 201 CACATTAGAACAATATCAGCTATGTGCTATGCTATATCAACGATGAT 250  
 916 yrGlnValArgGlnValPheAlaGlnLysLeuHisLysGly 929  
 251 ATCAAGTAAGACAGAGTGTTTGGCCCAAGAACTTCACAAAGGC 291

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA280598  
 seq\_documentation\_block:  
 ID AA280598 standard; cDNA; 530 BP.  
 XX  
 AC AA280598;  
 XX 07-APR-2000 (first entry)  
 XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:682.  
 XX Human; gene expression product; diagnosis; tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 XX Homo sapiens.  
 OS  
 XX WO9964576-A2.  
 XX 16-DEC-1999.  
 XX 09-JUN-1999; 99WO-IB01062.  
 XX 10-JUN-1998; 98US-0088801.  
 XX (FARB ) BAYER CORP.  
 PA Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Catino TU, Derti A, Ford DW, Lewis ME, Monahan JE;  
 PI Schlegel R;  
 XX WPI; 2000-087220/07.  
 DR Novel nucleic acids, used to develop products for the diagnosis and  
 XX treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer -  
 PT  
 PS Claim 15; Page 402; 469pp; English.  
 XX  
 CC AA279917 to AA280766 represent double stranded cDNA clones isolated from  
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g.  
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
 CC can be used for developing agents for the diagnosis and treatment of  
 CC disorders involving unwanted cell proliferation, such as neoplasia,  
 CC dysplasia or hyperplasia.  
 XX Sequence 530 BP; 191 A; 90 C; 83 G; 166 T; 0 other;  
 SQ  
 alignment\_scores:  
 Quality: 349.00 Length: 68  
 Ratio: 5.209 Gaps: 0  
 Percent Similarity: 98.529 Percent Identity: 98.529  
 alignment\_block:  
 US-09-512-581-2 x AA280598/rev ..  
 Align seg 1/1 to reverse of: AA280598 from: 1 to: 530  
 37 MetValLysThrPheMetAspMetAspGlnAspSerGluGluGly 53  
 213 ATGTTCTGTAACACTTTTATGGATATGCACGAGACTCTGAAGAAGAAA 164  
 53 sGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeu 70  
 163 GGAGCTTTATTAAACCTAGCTTTACATCTGCATCTGCTTCCTCA 114  
 70 yHisProGlyLysAspValArgLeuValAlaCysCysLeuAlaAsp 86







XX	12-OCT-1999	(first entry)
DT		
XX		
DE	Human gene expression product	cDNA sequence SEQ ID NO:2728.
DE		
KW	Human; gene;	gene expression product; diagnosis; therapy; probe;
KW	detection; mapping;	tissue typing; profiling; forensic; cancer;
KW	genetic analysis;	colorectal cancer; breast cancer; lung cancer; ss.
XX		
XX	Homo sapiens.	
OS		
XX	WO9338972-A2.	
PN		
XX		
PD	05-AUG-1999.	
XX		
PF	28-JAN-1999;	99WO-US01619.
XX		
PR	03-APR-1998;	98US-0080666.
PR	28-JAN-1998;	98US-0072910.

PK	31-MAR-1998;	98US
PR	03-APR-1998;	98US-
XX		
PA	(CHIR ) CHIRON CORP.	
PA	(HYSE-) HYSEQ INC.	

PI Crkvenjakov R, Dick  
PI Escobedo J, Garcia

PI Jones WL, Kassam A, Kennedy SC, Rice D, Labat I, Leshkowitz D, Pot D, Randazzo F, Reinhard C, Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C

PI Stacne-Crain B, Sudduth-Kilinger J, Williams LT;  
XX  
XX WPI; 1999-494092/41.  
DR  
XX

PS Claim 1: Page 1324: 2479pp: English.  
XX differentially expressed in different cell types  
PT novel human genes and their expression products which are

xx  
CC The present invention describes a library of human polynuc  
CC comprising the sequences given in AA212532 to AA217779. A  
CC a method of detecting differentially expressed genes corre  
CC

CC differentially expressed gene product in a test sample from  
CC cancerous state of a mammalian cell, comprising detecting  
CC differentially expressed gene product in a test sample from  
CC suspected of being cancerous, where the gene product is  
CC of the 5248 polynucleotide sequences given in AAZ12532 to

polynucleotides can be used as a source of primers and probes for a variety of purposes, e.g. detection of expression, mapping, tissue typing or profiling, forensics, genetic analysis of polymorphisms. Polynucleotides encoded by the nucleotide sequence of the present invention may also be used for

can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used as arrays for diagnostics (which may be used to determine function of encoded proteins) and to detect differences in expression of encoded proteins.

two cells (e.g. to identify abnormal or diseased tissue in a specimen). The polynucleotides of the invention are especially useful for identifying a genetic predisposition or susceptibility to a cancer. The polynucleotides of the invention are especially useful for the diagnosis, screening and management of colorectal cancer.

CC and lung cancer. The polynucleotides can also be used to  
CC peptide analogues and antagonists.

[illegible]

Ratio:	5.075	Gaps:	0
Percent Similarity:	98.361	Percent Identity:	95.902

```
alignment_block:
US-09-512-581-2 x AAZ15259 ..
```

Align seq 1/1 to: AAZ15259 from: 1 to: 738











478 TyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnG1 494  
152 TATTACTTATATAGTCTAGTTGGATCCAAATGCTGTAAAGCTCTCAACGA 201  
494 uMetTyrLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuA 511  
202 AATGTGGAGTGTGACAGAACTGCTCGGATCCATGTACGCGAATATGG 251  
511 sPLeuileLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527  
252 ATTTGCACAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGA 301  
528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAs 544  
302 AAACCTGATGACCATAGCAAGAAATTTGCGTGCACCGGGGAAGACACA 351  
544 pPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgL 561  
352 TTTTGTGAGAAATTTTACACAGGTTCTCGGCGATGATGAGAAATCTCG 401  
561 ySglnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGlu 577  
402 CTCAGTTGGAGTTTAAATAGCCCAACCTGTTCTTGCRAACAAGCAGAT 451  
578 GlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProTh 594  
452 ATTTGTGTGAGAAATAGCCCGAACTTGCAAATCCTTAAGCAACCAAC 501  
594 rAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProV 611  
502 AAATCTTTCTAGAGATGTCGCAATTTCTGTTGGAAAGAAATCGCACCT 551  
611 aHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627  
552 TGCACATGATTTCAGAGCCATAAGTGCACCTAGTGAATGATGAATAAG 601  
628 SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspG 644  
602 TCAATAGAGGGGACAGACATGATGAAGAGGAGGGTGTAAAGTCCAGATA 651  
644 nAlaIleArgAlaGlyLeuGluLeuLysValLeuSerPheThrHisP 661  
652 AGCTATCCGTTCCAGACTTGAACCTCTTAAGGTTCTCTTTTACACATC 701  
661 rotSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeu 677  
702 CTACCTCGTTCCACTCTGCAGACATATAGTCTTGTACAGTGCCTA 751  
678 LysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAs 694  
752 AGAATGGAGATGACAAAGGTAGCAGAGCTGCTATTCAAAATTTTAGAAA 801  
694 nThrGlySerLysIleGluAspPheProHisIleArgSerAlaLeuL 711  
802 TACAGGTCAAAATAGAAACACAGCTCTCCAGATACGATCCACCTTAA 851  
711 euProValLeuHisLysLysSerLysGlyProArgGlnAlaLys 727  
852 TTCCATTTTACATCAAAAGCAAGAGGGGTACTCCACACCAAGCAAAA 901  
728 TyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPh 744  
902 CAGGCTGTGCATGTATACAGCCCATATTCCAAATAAAGAAAGTCCAGCT 951  
744 eAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuG 761  
952 TGCACAGATTTTTCAGCCACTAGTAGGAGTCTGAATGCTGATGTCGCCAG 1001  
761 luHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAla 777  
1002 AACCACTTATACTCCATAGTTTTCATTGGGCCACATTTCTATGTAGCA 1051

778 ProAspGlnPheAlaAlaProThrLysSerTrpValAlaThrPheIleVa 794  
1052 CCAGATCAGTTTCTCCCAATGAAATCTGTAGTAGCAAAATTTATTGT 1101  
794 lLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrLysL 811  
1102 GAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAGAAATGGAANA 1151  
811 euTrpValProAspGluValSerProGluThrMetValLysIleGln 827  
1152 TGTGGTCTCCAGATGAAGAGTTTCCCTGAAGTACTAGCAAAAGGTACAG 1201  
828 AlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnHisSe 844  
1202 GCAATTAACCTTCTGTAAGGTGGCTGTTGGGTATGAAAAAACCAACG 1251  
844 rLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSerA 861  
1252 TAATCTGCCAATTAACCTTCCGTTATATCAGCGATGTTGGTTAGTG 1301  
861 sPglyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877  
1302 AGGTCACCTGCAGACAGCAAAAGAGGATCAGTAAATCTGATATGCTCGC 1351  
878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCy 894  
1352 TTGCGATTAGCTGCTGTAGTGCATATAATGAAGCTTGTCCAGGAACCTG 1401  
894 sTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAla 911  
1402 TTACCATGAATTTATACCCAGAACAGTTTTCAGCTCTGTCACATTGTTA 1451  
911 leAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHis 927  
1452 TTAATGATGAGTGTACCAGTAAGCAGATATTTGCTCAGAAGCTGCAT 1501  
928 LysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAl 944  
1502 AAGGCACCTTGTGAAGTTACTGCCATTTGGAGTATATGCGCATCTTTG 1551  
944 aLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnC 961  
1552 CTGTGTGCCAAAGATCTTGAAGGAGAGAGAGACAGCAGCAGCAGCAAT 1601  
961 ysLeuValLysAsnIleAsnValArgGluTyrLeuLysGlnHisAla 977  
1602 GTTTACTGAAAAATATCAGTATACGAGGGAATACATTAAAGCAGAACTCT 1651  
978 AlaValSerGluLysLeuSerLeuProGluTyrValValProTy 994  
1652 ATGGCTACTGAGAAATTTATATCAGTGTGCTGGAATATATGATGCTCCATA 1701  
994 rThrIleHisLeuAlaHisAspProAspTyrValLysValGlnAspI 1011  
1702 CATGATTACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAGATG 1751  
1011 leGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIle 1027  
1752 TTGATCAGCTTCTGTATCAAAAGAGTGCCTATGTTTCATGCTTGAAGTT 1801  
1028 LeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVa 1044  
1802 TTAATGACAAAGAAATGAAACCAATAGCCATGCTTTATGAAGAAGATGCC 1851  
1044 lGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAlaLysM 1061  
1852 AGAACAATCAAGTTAACAGAGATGCCAGTCTCCAGATGAATCCAAAGA 1901  
1061 etAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMet 1077  
1902 CAAATCAAAACCTGATACAGTATGATGTCGCTCTCTGCTGTTATAAAT 1951  
1078 SerLysSerThrThrSerLeuGluSerProLysAspProValLeuPr 1094



1267 uLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerP 1284  
 |||||  
 3950 AAAGAGGCTCAAGAGATATATTAGAAATGAAGATGAACAGAAATAGTC 3999  
 |||||  
 1284 rProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300  
 |||||  
 4000 CGCCAAAAAGGGTAAAGAGGCGCCACCACCAACCTCTTGGTGGAGGT 4049  
 |||||  
 1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317  
 |||||  
 4050 ACACCAAAAGAGAGCCCAACCAATGAACCTTCTAAAAAGAGCAAAAA 4099  
 |||||  
 1317 sLysSerGlyProProAlaProGluGluGluGluGluGluGlns 1334  
 |||||  
 4100 AAAATCTGGACCTCCAGCACGAGGAGGAGGAAGAGAGAGACAAA 4149  
 |||||  
 1334 eRGlyAsnThrGluGlnLysSerLysGlnHisArgValSerArg 1350  
 |||||  
 4150 GTGGAATACGGAACAGACAGTCCAAAGCAACACACCGAGTCTCAAGG 4199  
 |||||  
 1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367  
 |||||  
 4200 AGAGCAGACAGCAGAGAGATCTCTCGAATCTAGTGAATGAATCCAC 4249  
 |||||  
 1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP 1384  
 |||||  
 4250 ACATCCACACCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4299  
 |||||  
 1384 rOSerGlnProLysLysAsnVal 1391  
 |||||  
 4300 CATCAACCAACCAAAAAAATGTG 4322

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA47423

seq\_documentation\_block:

XX AAA47423 standard; DNA: 3957 BP.

XX AC AAA47423;

XX DT 20-OCT-2000 (first entry)

XX DE Sequence encoding human neuron-associated protein.

XX KW Neuron associated protein; NEUP; neurological disorder; epilepsy;  
 KW Ischemic cerebrovascular disease; stroke; cerebral neoplasm;  
 KW Alzheimer's disease; Pick's disease; Huntington's disease;  
 KW dementia; Parkinson's disease; demyelinating disease; meningitis;  
 KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;  
 KW cerebral palsy; muscular dystrophy; central nervous system; CNS;  
 KW peripheral nervous system; PNS; myopathy; schizophrenia;  
 KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;  
 KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;  
 KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;  
 KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;  
 KW AIDS; Addison's disease; adult respiratory distress syndrome;  
 KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;  
 KW Werner syndrome, trauma; human; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 140..2704

FT /\*tag= a

FT /product= Neuron associated protein

XX WO200034477-A2.

XX PN 15-JUN-2000.

XX PD 10-DEC-1999; 99WO-0530408.

XX PF 11-DEC-1998; 98US-0210083.

XX PR

PR 11-DEC-1998; 98US-9123456.  
 PR 09-FEB-1999; 99US-0119365.  
 PR 16-MAR-1999; 99US-0124687.  
 XX (INCY-).INCYTE PHARM INC.  
 XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;  
 PI Lu DAM, Azimzai Y;  
 XX WPI; 2000-423423/36.  
 DR P-PSDB; AAB01382.  
 XX New human neuron-associated proteins and polynucleotides encoding them,  
 PT useful for diagnosis, treatment and prevention of cell proliferative  
 PT disorders including cancer, neuronal and neurological disorders  
 XX Claim 9; Page 130-131; 145pp; English.  
 XX Human neuron-associated proteins (NEUP) can be used for for  
 CC treating or preventing a disorder associated with decreased  
 CC expression or activity of NEUP. Antagonists of NEUP are useful for  
 CC treating or preventing disorder associated with increased expression  
 CC or activity of NEUP. NEUP or their fragments or derivatives are  
 CC useful for treating neurological disorder such as epilepsy, ischemic  
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's  
 CC disease, Pick's disease, Huntington's disease, dementia and  
 CC Parkinson's disease. NEUPs are also useful for treating other  
 CC demyelinating diseases, bacterial and viral meningitis, prion  
 CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and  
 CC metabolic diseases of the nervous system, neurofibromatosis, other  
 CC developmental disorders of the central nervous system, cerebral  
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,  
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and  
 CC other neuromuscular disorders, peripheral nervous system disorders,  
 CC inherited, metabolic, endocrine, and toxic myopathies, mental  
 CC disorders including mood, anxiety and schizophrenic disorders, a cell  
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective  
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal  
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,  
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory  
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's  
 CC disease, adult respiratory distress syndrome, allergies, ankylosing  
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,  
 CC complications of cancer, hemodialysis, and extracorporeal circulation,  
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic  
 CC infections, and trauma. This sequence was given the Incyte ID no.  
 CC 2888437CB1.  
 XX SQ Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;

alignment\_scores:

Quality: 2932.50 Length: 977

Ratio: 3.693 Gaps: 16

Percent Similarity: 81.269 Percent Identity: 60.901

alignment\_block:

US-09-512-581-2 x AAA47423

Align seg 1/1 to: AAA47423 from: 1 to: 3957

428 AspAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTy 444

2 GAAGCTCAGAGAAAGTCAGCTGGATGAAGCAAACTTCTGCATATTTA 51

444 rTyrGlnAsnSerIleAspArgLeuLeuValGluArgIlePheAlaG 461

52 TTATCAGAACAGCATTTGACGACAACTGTTGGTAGAGAAAATCTTTCCTC 101

461 lntYrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeu 477

102 AGTATCTTGTCCCCACAACTCGGAAACAGAGAGAGAAATGAAATGCTTA 151



2100 GAACGTGCTTAAGGTACTCTCATTTTACACATCCCATTCATTTTCATTCTGC 2149  
 667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684  
 2150 TCAAAACATTGAATCATTTACTGGCTTGCTGAAATGGATGATGAAAAAG 2199  
 684 aAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700  
 2200 TAGCAGAAGCTGCATCTACAAATTTTCAAAACACAGAGAACAAAATGAA 2249  
 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLys 717  
 2250 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCCTGTTTACATCACAA 2299  
 717 sSerLysLysGlyProArgGlnAlaLysTyrAlaIleHisCysIleH 734  
 2300 ATCTAAAAAAGAGACCCCCCGCTCAAGGCAATATGCCATTCATTTGATTC 2349  
 734 iSAlaIlePheSerSerLysGlnThrGlnPheAlaGlnIlePheGluPro 750  
 2350 ATGCGATATTCTTAGTAAGAGACCAGTTTGACACAGATATTGTAGCCCT 2399  
 751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767  
 2400 CTGCATAAGACCTAGATCCAAAGCAACCTGGAACATCTCAFAACACCAT 2449  
 767 uValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaP 784  
 2450 GGTGTACTATTGGTCATATTGCTCTCTTGACCTGATCAATTTGCTGCTC 2499  
 784 rGTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800  
 2500 CTTGGAATCTGGGTAGCTACTTTCATTGTGAAGATCTTCTCATGCAAT 2549  
 801 AspArgLeuProGlyLysLysThrLysLeuTrpValProAspGluGl 817  
 2550 GATCGGCTTCAGGGAAAAAGACAACTAACTTTGGTGTCCAGATGAAGA 2599  
 817 uValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834  
 2600 AGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAATCATGGTTC 2649  
 834 rGTrpLeuGlyMetLysAsnHisSerLysSerGlyThrSerThr 850  
 2650 GATGGCTACTTGAATGAAAAATAATCACAGTAATCAGGAACCTTCTACC 2699  
 851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGl 867  
 2700 TTAAGATTGCTAACAACTATTGCATAGTGTGAGACTTGACAGAAC 2749  
 867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlyS 884  
 2750 GGGGAAATTAGTAAACCAGATATGTCAGCTCTGAGACTTGTCTGGGA 2799  
 884 eAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900  
 2800 GTGCTATTGTGAAGCTGGCACAGAACCCTGTTACCATGAATCAATCATCA 2849  
 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGl 917  
 2850 TTTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCA 2899  
 917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934  
 2900 AGTAAGACAAGTGTTTGGCCCAAGAACTTCACAAAGGCTTTTCGCCGTTTAC 2949  
 934 rGLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950  
 2950 GGCCTTCCACTTGAGTATTAGCAATCTGTGCCCTTTGTGCAGAAAGATCCT 2999  
 951 ValLysGluuArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967

3000	GTAAAGAGAGAGAGAGCTCATGCTAGGCAATGTTGGTGGAANAATATATAA	3049
967	nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL	984
3050	TGTAAGCGCGGAGTATCTGAAGCAGCATGCAGCTCTTACTGAAAAATTAT	3099
984	eusSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla	1000
3100	TGTCTCTTCTACCCAGAGATGTTGTTCATATACAAATTCACCTTTGGCA	3149
1001	HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa	1017
3150	CATGACCCAGATTATGTCAAAGTACAGCATATTGAACAACCTTAAAGATCT	3199
1017	LysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluA	1034
3200	TAAAGAAATGCTTTTGGTTGTCTGGAATATTAATGGCTAAAAATGAAA	3249
1034	snAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr	1050
3250	ATACAGTCACGCTTTTATCAGAAGATGTTAGAAAAATTTAAACAAACA	3299
1051	LysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrTh	1067
3300	AAAGATGCCCAAGGACCATGATGTCAAAAATGAATGAAAAACTGTACAC	3349
1067	rValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrS	1084
3350	TGTGTGTGATGTTGCCCATGAATATCATGATGCAAGAGTACTACATACA	3399
1084	erLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln	1100
3400	GTTTGGAACTCTCTAAAGACCGGTACTACCAGCTCGTTTCTTCACTCAA	3449
1101	ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLy	1117
3450	CCTGCACAAGAAATTCAGTAAACACCAAAAAATTATCTGCCTCTCGAATGAA	3499
1117	sSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaV	1134
3500	ATCATTTTTCACCTCTGGAAAACTTAAACAACCAATGTTCTTAGGAGCTG	3549
1134	alAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer	1150
3550	TTAAACAAGCCACTTTCATCAGCAGCAAGCAATCTCAGACCAAAATCATCA	3599
1151	ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerPr	1167
3600	CGAATGGAACTGTAAAGCAATGCAGCAGCAGCTCAATCCAAAGCTCTCC	3649
1167	oGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA	1184
3650	TGGAAGAATAAAGGGAGGCTGTAGTTCTCAAAATGGATCATCAGTGAAG	3699
1184	snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys	1200
3700	ATGAAGATTACAAATGTCTTCACCTTTGCCGGGAAAAAAGTGACAAG	3749
1201	ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyAr	1217
3750	AGAGACGACTCTGATCTTGTAAAGTCTCAATTGGAGAAAGCCTTAGAGCAG	3799
1217	gLysLysThrProValThrGluGlnGluLysLeuGlyMetAspAspL	1234
3800	GAAAAAACGCCCTCAGACAACAGGAGGAGAAATTAGGTATGGATGACT	3849
1234	euthrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg	1250
3850	TGACTAAGTTGGTACAGAAACAGAAACCTTAAAGGCATGTCAGCAAGTCCG	3899
1251	LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGl	1267
3900	AAAAGAGCCATACGGTTCAGAAATCTCATCAACAGCACTGCCCTGAGGA	3949



51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67  
|||||  
300 GAAGAAAGAGAGCTTTATTTAAACCTAGCTTTACATCTTCTCAGATTT 349  
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84  
|||||  
350 TTTTCTCAAGCATCTCGTGAAGATGTTCTGCTTACTGCTAGCTGCTGCC 399  
84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100  
|||||  
400 TTGCTGATATTTTCAGGATTTATGCTCTCTGAAGCTCTTACATCCCT 449  
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117  
|||||  
450 GATAAATCTAAGGATATATTTATGTTTATACAAAGCATGTTGAAGGGCT 499  
117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134  
|||||  
500 AGAGGATACAAAGAGCCACAAATTCATAGTATTTTATTTACTTGAGA 549  
134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150  
|||||  
550 ACATTTGCTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGC 599  
151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167  
|||||  
600 AATGAAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAA 649  
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184  
|||||  
650 TGGCCACAATCAGAAAGTCCATATGCACATGTTAGACCTTATGAGCTTA 699  
184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200  
|||||  
700 TTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTA 749  
201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217  
|||||  
750 GTAATCTGGTACCTGCTCATAGAAATTTAAACAAAGCAACATATGATTT 799  
217 uAlaLysAlaLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234  
|||||  
800 GGCAAGGCTTTTACTGAAGAGGAGCAGCTCAAGCTATTGAGCCATATATTA 849  
234 hrThrPheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250  
|||||  
850 CCACCTTTTAAATCAGGTTCTGATGCTTGGGAAACATCTATCAGCGAT 899  
251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267  
|||||  
900 TTGTCAGAGCATGCTTTTGACTTAATTTTGGAGCTCTACAAATATTGATAG 949  
267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284  
|||||  
950 TCATTTGCTGCTCTGTTTACCCAGCTTGAATTTAAATTAAGAGCA 999  
284 snAspAsnGluGluArgGlnValValLysLeuLeuAlaLysMetPhe 300  
|||||  
1000 ATGATAATAGGAGCGCTCAAGTGTGTTAACTTACTTGGCAAAATGTTT 1049  
301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317  
|||||  
1050 GGGCAAGGATTCAGAAATGGCTTCTCAAAACAAAGCCACTTTTGGCAGTG 1099  
317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334  
|||||  
1100 CTACTTGGCAGGTTTAAATGATATCCATGTACCAATCCGCGCTGGAATGTG 1149  
334 alLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350  
|||||  
1150 TGAATTTGTAGCATTTGTCTCATGAACCATCTCTGATTTAGCAAAAGAC 1199

351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367  
|||||  
1200 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTGAGGAAGCTATTAG 1249  
367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384  
|||||  
1250 ACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTC 1299  
384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400  
|||||  
1300 TGGTCAATGATCACTTACTTAAATTTGTGAGAGAGAACATTAGACAAA 1349  
401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417  
|||||  
1350 CGATGGAGAGTACGCAAGAGCCATCATGGGACTTCCCCAAATTTTATAA 1399  
417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434  
|||||  
1400 GAAATATGCTTTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAG 1449  
434 latrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450  
|||||  
1450 CATGGATCAAAAGACAAATTCGTACATATATATTATCAAAATAGTATTGAT 1499  
451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467  
|||||  
1500 GATCGACTACTTGTGAAACGGATCTTTGCTCAATACATGTTCTCTCACAA 1549  
467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484  
|||||  
1550 TTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACAC 1599  
484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500  
|||||  
1600 TGGATTTAAATGCTGTAAGACATGTAATGAAATGTGAAATGTCAAAAT 1649  
501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProLy 517  
|||||  
1650 CTGCTCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAAGCAACCCAA 1699  
517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534  
|||||  
1700 AACAGATGCCAGTCTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAA 1749  
534 rGsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550  
|||||  
1750 GAAATTTACCTGATCTCTGGTGAAGCTCAGGATTTTCATGAAGAAATTCACA 1799  
551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567  
|||||  
1800 CAGGTGTTAGAGATGATGAGAAATTAAGAAAGCGTTTAGAAGTACTTGT 1849  
567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584  
|||||  
1850 TAGTCCAACATGCTCTCTGCAAGCAGGCTGAAGGTTGTGCTGCTGAAATAA 1899  
584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600  
|||||  
1900 CTAAGAAGTTGGCAACCCCAACAGCTACAAATCTCTTCTGGAAATG 1949  
601 lLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSe 617  
|||||  
1950 ATCAAGTTTCTCTGGAGAGGATAGCACCTGTGCACATAGATACCGAATC 1999  
617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634  
|||||  
2000 TATCAGTGTCTTTATTAACAAGTGAACAAATCAATAGATGGAACAGCAG 2049  
634 spAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650  
|||||  
2050 ATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGACGAGGCTCT 2099  
651 GlutLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667



```

FT      exon      /number= 10
FT      /1207..1352
FT      /*tag= j
FT      /number= 11
FT      /1353..1504
FT      /*tag= k
FT      /number= 12
FT      /1505..1618
FT      /*tag= l
FT      /number= 13
FT      /1619..1700
FT      /*tag= m
FT      /number= 14
FT      /1701..1749
FT      /*tag= n
FT      /number= 15
FT      /1750..1889
FT      /*tag= o
FT      /number= 16
FT      /1890..2005
FT      /*tag= p
FT      /number= 17
FT      /2006..2111
FT      /*tag= q
FT      /number= 18
FT      /2112..2272
FT      /*tag= r
FT      /number= 19
FT      /2273..2396
FT      /*tag= s
FT      /number= 20
FT      /2397..2555
FT      /*tag= t
FT      /number= 21
FT      /2556..2624
FT      /*tag= u
FT      /number= 22
FT      /2625..2761
FT      /*tag= v
FT      /number= 23
FT      /2672..2885
FT      /*tag= w
FT      /number= 24
FT      /2886..3090
FT      /*tag= x
FT      /number= 25
FT      /3091..3205
FT      /*tag= y
FT      /number= 26
FT      /3206..3338
FT      /*tag= z
FT      /number= 27
FT      /3339..3458
FT      /*tag= aa
FT      /number= 28
FT      /3459..3521
FT      /*tag= ab
FT      /number= 29
FT      /3522..3667
FT      /*tag= ac
FT      /number= 30
FT      /3668..3773
FT      /*tag= ad
FT      /number= 31
FT      /3774..4213
FT      /*tag= ae
FT      /number= 32
FT      /4214..4438
FT      /*tag= af
FT      /number= 33
FT      /4314..4408
FT      misc_signal
FT      /*tag= ag
FT      /note= "Destabilizing signal"

```

```

FT      exon      4439..5337
FT      /*tag= ah
FT      /number= 34
FT      /5312..5317
FT      /*tag= ai
FT      /5333..5337
FT      /*tag= aj
FT      /note= "Poly adenylation cleavage site"
XX      WO200050454-A1.
XX      31-AUG-2000.
XX      24-FEB-2000; 2000WO-US04732.
XX      24-FEB-1999; 99US-0121461.
XX      (TUFT ) TUFTS COLLEGE.
XX      Soto AM, Sonnenschein C, Geck P, Szelei J;
XX      WPI; 2000-565451/52.
XX      P-PSDB; AAY94702.
XX      New human androgen-induced tumor suppressor cDNA sequence termed
XX      'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX      diagnosis and treatment of prostate cancer -
XX      Example 4; Fig 6; 152pp; English.
XX      This invention relates to a human androgen-induced tumour suppressor
XX      cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
XX      located on chromosome 13 at position 13q12-13q. AS3 has a role in
XX      inhibiting cell proliferation and use as a marker for the efficient
XX      diagnosis and treatment of prostate cancer. The invention includes AS3
XX      cDNA and protein sequences, a vector comprising the cDNA sequence, a host
XX      cell transfected with the expression vector, and a method for producing
XX      an AS3 polypeptide comprising culturing the transfected cells. AS3 has
XX      cytoskeletal activity, and acts to suppress cell proliferation. The AS3
XX      gene is useful as a marker for the efficient diagnosis and treatment of
XX      prostate cancer. The AS3 nucleic acid molecule can be used as a source of
XX      antisense agents for sequence specific modulation of gene expression. The
XX      AS3 protein may be used in the treatment of disorders caused by aberrant
XX      modification or mutation of a gene encoding an AS3 protein, misregulation
XX      of the AS3 gene or aberrant post-translational modification of the AS3
XX      protein. This sequence represents the human AS3 cDNA sequence with an
XX      additional 84 nucleotides in the 5' untranslated region (5' UTR) when
XX      compared with the claimed AS3 cDNA sequence AAA28051.
XX      Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

alignment_scores:
    Quality: 7193.00      Length: 1391
    Ratio: 5.171
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-512-581-2 x AAA28052      ..

Align seg 1/1 to: AAA28052 from: 1 to: 5355

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17
|||||
150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAAATTACATATCCGCC 199

17 oGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArg 34
|||||
200 TGGGGTCAAGGAAATATACGATAAAATATCTAAAGAGGAGATGGTGAGAC 249

34 rgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
|||||
250 GATTAAGATGGTGTGTGAAAACCTTTTATGGATATGGACCAGGACTCTGAA 299

```



3066	CATGACCCAGATTATGTCAAAGTACAGGATATTGAAACAACTTAAAGATGT	3115
1017	llysGluCyLeuThrPheValLeuGluIleuMetAlaIysAsnGluA	1034
3116	TAAAGAATGCTTTTGTTGTTCTGGAAATATTAAATGGCTAAAAATGAAA	3165
1034	snaSserHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr	1050
3166	ATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAAATATTAACAACA	3215
1051	LysAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyrTh	1067
3216	AAAGATGCCCAAGGACAGATGATGCAAAAATGAATGAAAACGTGTACAC	3265
1067	rValCyAspValAlaMetAsnIleIleMetSerLysSerThrTyrS	1084
3266	TGTGTGTGATGTGGCCATGAATATCATGTCAAGAGACTACATACA	3315
1084	erLeuGluSerProLysAspProValLeuProAlaIleArgPheThrGln	1100
3316	GTTTGGAAATCTCCTAAAGACCCGGTACTACAGCTCGTTTCTTCACTCAA	3365
1101	ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLys	1117
3366	CCTGCACAGAATTTTCAGTAACACCAAAAATTTATCTGCCTCTCGAAATGAA	3415
1117	sSerPheThrProGlyLysProLysThrThrAsnValLeuGluValay	1134
3416	ATCATTTTTCACCTCTGAAACCTTAAACAACCAATGTTCTAGAGCTG	3465
1134	alaSnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer	1150
3466	TTAACAAGCCACTTTCATCAGCAGCAAGCAATCTCAGACCAATCATCA	3515
1151	ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerPr	1167
3516	CGAATGGAAACTGTAAAGCAATGCAGCAGCAGCTCAATCCAAAGCTCTCC	3565
1167	oGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA	1184
3566	TGAGAGATTAAGGGGAGGCTGTAGTGTCTGAAATGGATCAGTCAAGAAA	3615
1184	snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys	1200
3616	ATCAAGATTACAAATGCTTCACCTTTGCGCGGAAAAAAGTGACAAG	3665
1201	ArgAspASerAspLeuValArgSerGluLeuGluLysProArgGlyAr	1217
3666	AGAGACGACTGTGATCTGTAGGTCTGAAATGGAGAAGCCTAGAGGCAG	3715
1217	glyLysThrProValThrGluGlnGluLysLeuGlyMetAspAspL	1234
3716	GAAAAAACGCCGTCACAGAACAGGAGGAGAAATTAGTATGGATGACT	3765
1234	euThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArg	1250
3766	TGACTTAAGTTGTTACAGAACAGAAACCTAAAGCGACTACCGAAGTCGG	3815
1251	LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnThrProGluGl	1267
3816	AAAAGAGGCCATACGGCTTCAGATCTCATCAACAGCAGTGGCTTGAGGA	3865
1267	uLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerP	1284
3866	AAAGAGGCTCAAGAAGATATATTAGAAAAATGAAGATGAACAGAAATAGTC	3915
1284	roProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly	1300
3916	CGCCAAAAAGGGTAAAGAGGCCGACCCACCAAACTCTTGTGTGGAGGT	3965
1301	ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy	1317

3966	ACACCAAAAGAGAGCCACCATGAAAACTTCTATAAAAAGGAAGCACA AAAA	4010
1317	slysserGlyProProAlaProGluGluGluGluGluGluArgGlns	1334
4016	AAAAATCTGGACCTCCAGCACACCAGAGCAGAGGAAGAAGAAGAACAAA	4065
1334	erGlyAsnThrGluGlnLysSerLysSerLysGlnHisargValSerArg	1350
4066	GTGGAAATACGGAACAGAAGTCACAAAGCAAAACAGCACCGAGTGTC AAGG	4115
1351	ArqAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh	1367
4116	AGAGCACAGCAGAGACGAATCTCTGAATCTAGTGCAATTGAAATCCAC	4165
1367	rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP	1384
4166	ACAGTCCACACACAGAAAGGACGAGAGACCATCAAAACGCCATCAC	4215
1384	roSerGlnProLysLysAsnVal	1391
4216	CATCACACCAAAAAAAAAAATGTG	4238
seq_name: /SIDSl/cgcdata/geneseq/geneseqn/NA2000.DAT:AAA28052		
seq_documentation_block:		
ID	AAA28052 standard; cDNA: 5355 BP.	
XX	AA	AAA28052;
DT	01-DEC-2000 (first entry)	
XX	Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4	
XX	Androgen-induced tumour suppressor; androgen shutoff gene 3	
KW	Chromosome 15q12-13q; cell proliferation inhibitor; prostate	
KW	diagnosis; treatment; cytostatic; human; ss.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
FH	exon	85..130
FT		/tag= a
FT		/number= 1
FT	exon	131..257
FT		/tag= b
FT		/number= 2
FT		150..4325
FT		/tag= c
FT		/product= "AS3"
FT		/note= "Androgen shutoff gene 3 protein"
FT	exon	258..461
FT		/tag= d
FT		/number= 3
FT	exon	462..548
FT		/tag= e
FT		/number= 4
FT	exon	549..646
FT		/tag= f
FT		/number= 5
FT	exon	647..773
FT		/tag= e
FT		/number= 6
FT	exon	774..938
FT		/tag= f
FT		/number= 7
FT	exon	939..995
FT		/tag= g
FT		/number= 8
FT	exon	996..1111
FT		/tag= h
FT		/number= 9
FT	exon	1112..1206
FT		/tag= i



401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417  
 |||||  
 1266 CGATGGAGAGTAGCGCAAGAAGCCATGATGGGACTTGGCCAAATTTATAA 1315  
 |||||  
 417 slySTyRAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434  
 |||||  
 1316 GAAATATGCTTTACATGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAG 1365  
 |||||  
 434 latrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450  
 |||||  
 1366 CATGGATCAAGACAAATTCGTACATATATATCAAAATAGTATTGAT 1415  
 |||||  
 451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467  
 |||||  
 1416 GATCGACTACTTGTGAACGGATCTTTGTGCTCAATACATGTTCTCCACAA 1465  
 |||||  
 467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484  
 |||||  
 1466 TTTAGAACTACAGAACGGATGAATGCTTATATTACTTGTATGCCACAC 1515  
 |||||  
 484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500  
 |||||  
 1516 TGGATTTAAATGCTGTGAAAGCATTCGAATGAAATGTGAAATGTCAAAAT 1565  
 |||||  
 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeuIleLysGlnProLy 517  
 |||||  
 1566 CTGCTCCGACATCAAGTAAGGATTTGCTTGACTTGATTAAGCAACCCAA 1615  
 |||||  
 517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534  
 |||||  
 1616 ACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGTGGTTATTACAA 1665  
 |||||  
 534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550  
 |||||  
 1666 GAAATTTACCTGATCTCTGTGTAAGGCTCAGGATTTTCATGAAGAAATTCACA 1715  
 |||||  
 551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567  
 |||||  
 1716 CAGGTGTTAGAAGATGATGAGAAAATTAAGAAAGCAGTTAGAAGTACTTGT 1765  
 |||||  
 567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584  
 |||||  
 1766 TAGTCCAAACATGCTCTCGAACAGGCTGAAGGTTGTGCGTGAATAA 1815  
 |||||  
 584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600  
 |||||  
 1816 CTAAGAAGTTGGCAACCCCAACAGCCCTACAAATCCTTTCTGGAAATG 1865  
 |||||  
 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluse 617  
 |||||  
 1866 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCATAGATACCGAATC 1915  
 |||||  
 617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634  
 |||||  
 1916 TATCAGTGTCTTATTAACAGTGAACAAATCAATAGATGAACAGCAG 1965  
 |||||  
 634 sPAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650  
 |||||  
 1966 ATGATGAAGATGAGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTT 2015  
 |||||  
 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667  
 |||||  
 2016 GAACGTCTTAAGGTACTCTCATTTACACATCCCATCTCATTTTCATTTCTGC 2065  
 |||||  
 667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684  
 |||||  
 2066 TGAACATTTGAATCATTTACTGGCTTGTCTGAAAATGGATGATGAAAAAG 2115  
 |||||  
 684 alalaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700  
 |||||  
 2116 TAGCAGAAGCTGCATACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2165  
 |||||

701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLy 717  
 |||||  
 2166 GAGGATTTTCCACACATCAGATCAGCCTTGTCTTCTCTTTTACATCACA 2215  
 |||||  
 717 sSerLysLysGlyProArgGlnAlaLysTyrAlaIleHisCysIleH 734  
 |||||  
 2216 ATCTAAAAAAGGACCCCGCTCAAGCAAAATATGCCATTCATTGTGTATCC 2265  
 |||||  
 734 isAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750  
 |||||  
 2266 ATGCGATATTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCT 2315  
 |||||  
 751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767  
 |||||  
 2316 CTGCATTAAGAGCCTAGATCCAAGCAACCTCGAACAATCTCATACACCAT 2365  
 |||||  
 767 uValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlap 784  
 |||||  
 2366 GGTACTATTGGTCATATTGCTCTCTTGCACCTGATCAATTTGCTGCTC 2415  
 |||||  
 784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800  
 |||||  
 2416 CTGTGAAATCTTGGGTAGCTACTTTCATTTGTGAAGATCTTCTCATGAAT 2465  
 |||||  
 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluG 817  
 |||||  
 2466 GATCGGCTTCCAGGGAAAAAGACAACATAAACTTTGGGTTCCAGATGAAGA 2515  
 |||||  
 817 uValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834  
 |||||  
 2516 AGTATCTCTGAGACAAATGTCAAAATTCAGGCTATTAAATGATGGTTC 2565  
 |||||  
 834 rGTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850  
 |||||  
 2566 GATGGCTACTTGAATGAAAAATAATACACAGTAAATCAGGAACCTTCTACC 2615  
 |||||  
 851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluG 867  
 |||||  
 2616 TTAAGATTCTTAACAACAATATTGCTAGTAGTGGAGACTTGACAGAACA 2665  
 |||||  
 867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlyS 884  
 |||||  
 2666 GGGGAAAAATTAGTAAACAGATATGTCACGTCTGACACTTGTCTGGGA 2715  
 |||||  
 884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900  
 |||||  
 2716 GTGCTATTGTGAAGCTGGCACAGAACCCTGTTTACCATTGAAATCATACA 2765  
 |||||  
 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG 917  
 |||||  
 2766 TTAGAACAATATCAGCTATGTGCATTTAGCTATCAACGATGAATGCTATCA 2815  
 |||||  
 917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934  
 |||||  
 2816 AGTAAGACAAGTGTGGCCCAAGAACTTCACAAAGGCCCTTCCCGTTTAC 2865  
 |||||  
 934 rGLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950  
 |||||  
 2866 GGCCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCT 2915  
 |||||  
 951 ValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967  
 |||||  
 2916 GTAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTGGTGAATAATATAA 2965  
 |||||  
 967 nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL 984  
 |||||  
 2966 TGTAAGCGGGAGATCTCAAGCAGCATGCGCTGTTAGTGAATAATATAT 3015  
 |||||  
 984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000  
 |||||  
 3016 TGTCTCTTCTACAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065  
 |||||  
 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysaspVa 1017  
 |||||



PN WO200050454-A1.  
 XX 31-AUG-2000.  
 XX 24-FEB-2000; 2000WO-US04732.  
 XX 24-FEB-1999; 99US-0121461.  
 XX (TUFT) TUFTS COLLEGE.  
 XX Soto AM, Sonnenschein C, Geck P, Szelei J;  
 XX WPI; 2000-565451/52.  
 DR P-PSDB; AAY94702.  
 XX New human androgen-induced tumor suppressor cDNA sequence termed  
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient  
 PT diagnosis and treatment of prostate cancer -  
 XX Claim 1; Fig 1; 152pp; English.  
 XX This invention relates to a human androgen-induced tumour suppressor cDNA  
 CC sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located  
 CC on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell  
 CC proliferation and use as a marker for the efficient diagnosis and  
 CC treatment of prostate cancer. The invention includes AS3 cDNA and protein  
 CC sequences, a vector comprising the cDNA sequence, a host cell transfected  
 CC with the expression vector, and a method for producing an AS3 polypeptide  
 CC comprising culturing the transfected cells. AS3 has cytostatic activity,  
 CC and acts to suppress cell proliferation. The AS3 gene is useful as a  
 CC marker for the efficient diagnosis and treatment of prostate cancer. The  
 CC AS3 nucleic acid molecule can be used as a source of antisense agents for  
 CC sequence specific modulation of gene expression. The AS3 protein may be  
 CC used in the treatment of disorders caused by aberrant modification or  
 CC mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene  
 CC or aberrant post-translational modification of the AS3 protein. This  
 CC sequence represents the human AS3 cDNA sequence.  
 XX  
 SQ Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;

## alignment\_scores:

Quality: 7193.00 Length: 1391  
 Ratio: 5.171 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-512-581-2 x AAA28051 ..

Align seg 1/1 to: AAA28051 from: 1 to: 5271

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17  
 66 ATGGCTCATTTCAAAGACTAGGACCAATCATGTGAAATAATACATATCCGCC 115  
 17 oGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgA 34  
 116 TGGGGTCAAGGAATATCAGATAAAATATCTTAAAGAGAGATGTGTGAGAC 165  
 34 rgLeuLysMetValValLysThrPheMetAspMetAspGlnAnSpSerGlu 50  
 166 GATTAAGATGGTTGTGAAACTTTTATGGATATGGACGAGGACTCTGAA 215  
 51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67  
 216 GAAGAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCCTCAGATTT 265  
 67 ePheLeuLysHisProGlyLysAspValArgLeuValAlaCysCysL 84  
 266 TTTTCTCAAGCATCTCGTAAAGATGTTCGCTTACTGGTAGCCCTGCTGCC 315  
 84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100

316 TTGCTGATATTTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCT 365  
 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117  
 366 GATAAACTAAAGGATATATTTATGTTTATACAAAGACAGTTGAAGGGCT 415  
 117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134  
 416 AGAGGATACAAAGAGCCCAATTCATAGGTATTTTATTTACTTGAGA 465  
 134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150  
 466 ACATTTGCTTGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGATAGC 515  
 151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167  
 516 AATGAAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATAACAA 565  
 167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerI 184  
 566 TGGCCACAATCAGAAAGTCCATATGCATGCTAGAGCTTATGAGCTCTA 615  
 184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200  
 616 TTATTTCTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGTTTA 665  
 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217  
 666 GTAATCTGTGATCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTT 715  
 217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234  
 716 GGCNAAGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTA 765  
 234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250  
 766 CCACCTTTTTTAAATCAGTCTGATGCTTGGGAAACATCTATCAGCGAT 815  
 251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267  
 816 TTGTGACAGCATGCTTTTGACTTAAATTTGGAGCTCTACAATATGTATAG 865  
 267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284  
 866 TCATTTGCTGCTCTCTGTTTACCCAGCTTGAATTTAAATTAAGAGCA 915  
 284 snAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300  
 916 ATGATAATGAGGAGCGCTACAAGTTGTTAAACTACTTGGCAAAAATGTTT 965  
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317  
 966 GGGCAAGAGATTTCAGAAATTTGGCTTCTCAAACAAGCCACTTTGGCAGTG 1015  
 317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334  
 1016 CTACTTGGGAGAGTTTAAATGATATCCATGTACCAATCCGCTCGAATGTG 1065  
 334 alLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysasp 350  
 1066 TGAATTTTGCTAGCCATTGCTCATGAACCATCTGATTTTAGCAAAAGAC 1115  
 351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367  
 1116 TTAACAGAGTATCTTAAAGTAGTGTCATGACCTGAGGAGGATATTAG 1165  
 367 qHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384  
 1166 ACATGATGTTATTGTGTCAATAGTTACAGCTGCTTAAAGAGGATATCTTC 1215  
 384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400  
 1216 TGGTCAATGATCACTTACTTAAATTTTGTGTGAGAGAGAGAACATTAGACAA 1265



```

/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT72157 + 129.00 142.91 5.45 2126
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAD00750 + 129.00 135.99 13.24 4193
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF54895 + 129.00 135.53 14.04 4386
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT05406 + 129.00 135.33 14.40 4472
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT16231 + 129.00 133.69 17.78 5254
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAO79377 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV82916 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA95039 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAZ38730 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC90510 + 129.00 130.08 28.27 7492
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV25812 + 129.00 130.08 28.27 7493

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28051
seq_documentation_block:
ID AAA28051 standard; cDNA; 5271 BP.
XX
AC AAA28051;
XX
DT 01-DEC-2000 (first entry)
XX
XX Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW chromosome 19q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.
XX
OS Homo sapiens.
XX
FH key
FT exon Location/Qualifiers
FT 1..46
FT /*tag= a
FT /number= 1
FT 47..173
FT /*tag= b
FT /number= 2
FT 66..4241
FT /*tag= c
FT /product= "AS3"
FT /note= "Androgen shutoff gene 3 protein, the CDS is
FT specifically claimed as SEQ ID #3"
FT
FT exon 174..377
FT /*tag= d
FT /number= 3
FT 378..464
FT /*tag= e
FT /number= 4
FT 465..562
FT /*tag= f
FT /number= 5
FT 563..689
FT /*tag= e
FT /number= 6
FT 690..854
FT /*tag= f
FT /number= 7
FT 855..911
FT /*tag= g
FT /number= 8
FT 912..1027
FT /*tag= h
FT /number= 9
FT 1028..1122
FT /*tag= i
FT /number= 10
FT 1123..1268
FT /*tag= j
FT /number= 11
FT 1269..1420
FT /*tag= k
FT /number= 12
FT 1421..1534
FT /*tag= l
FT
FT exon /number= 13
FT 1535..1616
FT /*tag= m
FT /number= 14
FT 1617..1665
FT /*tag= n
FT /number= 15
FT 1666..1805
FT /*tag= o
FT /number= 16
FT 1806..1921
FT /*tag= p
FT /number= 17
FT 1922..2027
FT /*tag= q
FT /number= 18
FT 2028..2188
FT /*tag= r
FT /number= 19
FT 2189..2312
FT /*tag= s
FT /number= 20
FT 2313..2471
FT /*tag= t
FT /number= 21
FT 2472..2540
FT /*tag= u
FT /number= 22
FT 2541..2677
FT /*tag= v
FT /number= 23
FT 2678..2801
FT /*tag= w
FT /number= 24
FT 2802..3006
FT /*tag= x
FT /number= 25
FT 3007..3121
FT /*tag= y
FT /number= 26
FT 3122..3254
FT /*tag= z
FT /number= 27
FT 3255..3374
FT /*tag= aa
FT /number= 28
FT 3375..3437
FT /*tag= ab
FT /number= 29
FT 3438..3583
FT /*tag= ac
FT /number= 30
FT 3584..3689
FT /*tag= ad
FT /number= 31
FT 3690..4129
FT /*tag= ae
FT /number= 32
FT 4130..4354
FT /*tag= af
FT /number= 33
FT 4320..4324
FT /*tag= ag
FT /note= "Destabilizing signal"
FT 4355..5253
FT /*tag= ah
FT /number= 34
FT 5228..5233
FT /*tag= ai
FT 5249..5253
FT /*tag= aj
FT /note= "Poly adenylation cleavage site"
FT
FT XX

```



/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAC51256	132.00	140.12	7.79	4040	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC51256	130.00	141.96	6.16	2640
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC51257	132.00	138.97	9.03	4522	/SDSI/cgdata/geneseq/geneseq/NA1995.DAT.AAC51257	130.00	139.81	8.11	3259
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA15096	132.00	138.02	10.21	4967	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA15096	130.00	139.39	6.59	3405
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAV02301	132.00	137.80	10.50	5074	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV02301	130.00	138.71	9.35	3632
/SDSI/cgdata/geneseq/geneseq/NA1990.DAT.AAO05831	132.00	133.75	17.66	7555	/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAV05831	130.00	136.31	12.71	4596
/SDSI/cgdata/geneseq/geneseq/NA1989.DAT.AAN91779	132.00	131.39	23.89	9520	/SDSI/cgdata/geneseq/geneseq/NA1992.DAT.AAO27189	130.00	136.14	12.99	4673
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT45424	132.00	131.28	24.22	9620	/SDSI/cgdata/geneseq/geneseq/NA1994.DAT.AAC66615	130.00	131.94	22.26	7056
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAT36211	132.00	128.06	36.61	13198	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC676473	130.00	130.09	28.21	3458
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAT36212	132.00	125.71	49.48	16622	/SDSI/cgdata/geneseq/geneseq/NA1992.DAT.AAO20610	130.00	129.45	30.32	3937
/SDSI/cgdata/geneseq/geneseq/NA1996.DAT.AAT06769	132.00	120.18	100.55	28598	/SDSI/cgdata/geneseq/geneseq/NA1993.DAT.AAO206185	130.00	129.53	30.64	3009
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT89956	132.00	120.06	102.20	28958	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV25810	130.00	129.45	30.64	3009
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAT75299	132.00	120.06	102.20	28958	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV25810	130.00	129.45	30.64	3009
/SDSI/cgdata/geneseq/geneseq/NA1993.DAT.AAO46806	132.00	119.74	106.47	29879	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX91162	130.00	129.45	30.64	3009
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAT06806	132.00	115.87	174.85	43676	/SDSI/cgdata/geneseq/geneseq/NA2001.DAT.AAC90508	130.00	129.45	30.64	3009
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAT06769	132.00	114.62	205.125	49377	/SDSI/cgdata/geneseq/geneseq/NA1994.DAT.AAO55134	130.00	128.18	36.06	1020
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAT68070	131.50	145.50	3.91	2241	/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT46941	130.00	127.69	38.39	1070
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA28521	131.50	145.50	3.93	2250	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX82361	130.00	126.66	43.81	1184
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX20288	131.50	142.56	5.70	2991	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV18277	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA14958	131.50	142.45	5.78	3024	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV18279	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT97611	131.50	140.58	7.35	3633	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAT22911	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA1994.DAT.AAO58708	131.50	140.32	7.60	3728	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX32569	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA1989.DAT.AAN90489	131.50	137.21	11.32	5056	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX35271	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC87968	131.50	136.90	11.79	5214	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX35271	130.00	124.11	60.78	1521
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA14824	131.50	136.90	11.79	5214	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX14850	130.00	107.52	93.30	2112
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT68770	131.50	131.28	24.21	9046	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA14851	130.00	120.77	51.16	7753
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT68761	131.50	131.28	24.21	9046	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX89807	129.50	154.85	1.18	700
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA81502	131.50	111.10	322.56	65632	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX39869	129.50	154.07	1.30	756
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAT21544	131.50	94.04	2.9e+03	349980	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC53422	129.50	149.64	2.30	1168
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA81490	131.50	79.65	1.7e+04	1437668	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC53422	129.50	149.64	2.30	1168
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA6314	131.00	151.78	1.85	1138	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC34615	129.50	147.50	3.02	1440
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC69638	131.00	148.84	2.24	1377	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC34945	129.50	146.74	3.33	1552
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAT231963	131.00	148.37	2.71	1590	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC48814	129.50	146.74	3.34	1553
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX83421	131.00	145.53	3.89	2101	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC48807	129.50	146.74	3.40	1574
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAX50286	131.00	145.53	3.89	2101	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC47915	129.50	146.28	3.54	1624
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAX83420	131.00	145.26	4.03	2158	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC49796	129.50	146.28	3.54	1624
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA50285	131.00	144.26	4.03	2158	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC49796	129.50	146.09	3.62	1654
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV42798	131.00	144.45	4.47	2336	/SDSI/cgdata/geneseq/geneseq/NA1995.DAT.AAC2645	129.50	146.07	3.63	1686
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC75869	131.00	142.50	5.75	2830	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA72650	129.50	142.61	5.66	2328
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT72156	131.00	142.44	5.79	2846	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA70260	129.50	139.88	8.05	3045
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAT78193	131.00	141.77	6.31	3040	/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT84930	129.50	139.65	8.28	3114
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AA241354	131.00	141.70	6.37	3061	/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AAT84930	129.50	139.65	8.28	3114
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA57868	131.00	141.66	6.40	3073	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV22451	129.50	139.06	8.94	3300
/SDSI/cgdata/geneseq/geneseq/NA2001.DAT.AAF58616	131.00	139.79	8.13	3691	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV17259	129.50	138.48	9.62	3492
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT97610	131.00	135.35	8.60	3854	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV11348	129.50	138.48	9.62	3492
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC51553	131.00	138.49	8.96	3975	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAC936459	129.50	138.20	9.98	3591
/SDSI/cgdata/geneseq/geneseq/NA1995.DAT.AAO82834	131.00	139.08	9.61	4193	/SDSI/cgdata/geneseq/geneseq/NA1989.DAT.AAN91475	129.50	138.07	10.15	3637
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AA237097	131.00	137.96	10.28	4416	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAV11344	129.50	136.60	12.25	4200
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AA245329	131.00	136.43	12.52	5136	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AA245329	129.50	136.60	12.25	4200
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT861548	131.00	134.08	16.91	6463	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AA248567	129.50	136.12	13.02	4402
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT861548	131.00	133.58	18.04	6791	/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AA248567	129.50	133.82	17.01	5520
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT861548	131.00	130.70	26.10	9009	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA60104	129.50	133.19	18.96	5868
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT85320	131.00	129.44	30.67	10195	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA60123	129.50	133.19	18.96	5868
/SDSI/cgdata/geneseq/geneseq/NA1997.DAT.AAT85320	131.00	127.27	40.52	12616	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA60123	129.50	133.19	18.96	5868
/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA88744	131.00	125.36	51.80	15225	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA60124	129.50	133.15	24.02	7032
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAT18276	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA60124	129.50	131.31	24.02	7032
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AA222910	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAD009831	129.50	131.09	24.83	7215
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AA222910	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1995.DAT.AAD09456	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1999.DAT.AA222910	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX35268	131.00	129.10	89.10	534720	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA2000.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAA62948	129.50	130.93	25.34	7326
/SDSI/cgdata/geneseq/geneseq/NA1998.DAT.AAX30458	131.00	125.35	51.82	15229	/				







/SIDS1/gcgdata/geneseq/NA1999.DAT.AAX07348 +	141.50	154.74	1.20	3089	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC76340 +	138.00	146.90	3.27	4341
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14691 +	141.50	151.06	1.92	4433	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14683 +	138.00	146.73	3.34	4414
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA93361 +	141.50	150.18	2.15	4835	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA09325 +	138.00	146.17	3.59	4660
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA27637 +	141.50	150.18	2.15	4833	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA742751 +	138.00	142.61	5.66	6008
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA21518 +	141.50	148.57	2.64	5661	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA18884 +	138.00	139.46	8.49	9009
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA72031 +	141.50	148.49	2.67	5707	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA76857 +	138.00	139.37	8.59	9088
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA70074 +	141.50	148.49	2.67	5707	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA59470 +	138.00	133.92	17.27	1551
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA60570 +	141.50	148.49	2.67	5707	/SIDS1/gcgdata/geneseq/NA1993.DAT.AAA040480 +	138.00	133.67	17.82	1589
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA78019 +	141.50	148.20	2.77	5968	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA73805 +	138.00	126.48	44.85	3220
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA17482 +	141.50	148.13	2.79	5910	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA72674 +	138.00	121.93	80.39	5034
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA28018 +	141.50	146.53	3.43	6916	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA39519 +	138.00	121.93	80.39	5034
/SIDS1/gcgdata/geneseq/NA1989.DAT.AAA95403 +	141.00	162.78	0.4266	1320	/SIDS1/gcgdata/geneseq/NA1995.DAT.AAA751411 +	138.00	121.54	84.50	5229
/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA84595 +	141.00	157.94	0.7927	2121	/SIDS1/gcgdata/geneseq/NA1993.DAT.AAA047357 +	138.00	121.54	84.50	5229
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA76442 +	141.00	154.14	1.29	3081	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA19941 +	138.00	111.69	298.84	1375
/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA71719 +	141.00	150.45	2.07	4424	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA49804 +	137.50	158.10	0.7769	1359
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA1468 +	141.00	150.42	2.08	4438	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA70096 +	137.50	154.34	1.26	1865
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA32553 +	141.00	150.30	2.11	4492	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA70212 +	137.50	154.18	1.29	1998
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA07840 +	141.00	143.28	5.20	8949	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA726543 +	137.50	153.68	1.37	2097
/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA28060 +	141.00	141.80	6.29	10347	/SIDS1/gcgdata/geneseq/NA1995.DAT.AAA080908 +	137.50	153.09	1.48	2223
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA26690 +	141.00	127.10	41.43	43804	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA56719 +	137.50	152.61	1.57	2329
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC76077 +	140.50	163.75	0.3764	1128	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA561632 +	137.50	149.20	2.43	3257
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC51066 +	140.50	160.70	0.5567	1522	/SIDS1/gcgdata/geneseq/NA1995.DAT.AAQ91432 +	137.50	148.05	2.82	3645
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA91902 +	140.50	150.86	1.97	4000	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA59395 +	137.50	148.05	2.82	3645
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA076173 +	140.50	148.82	2.56	4886	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA40483 +	137.50	147.63	2.98	3798
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA71738 +	140.50	147.23	3.13	5709	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14679 +	137.50	146.09	3.62	4417
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA17467 +	140.50	146.87	3.28	5913	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAT59909 +	137.50	145.14	4.09	4849
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA68385 +	140.50	146.84	3.29	5934	/SIDS1/gcgdata/geneseq/NA2000.DAT.AA287101 +	137.50	144.81	4.27	5008
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA33181 +	140.50	145.69	3.82	6644	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA75237 +	137.50	143.16	5.28	5893
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA33182 +	140.50	144.63	4.37	7372	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA59979 +	137.50	143.16	5.28	5893
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA33180 +	140.50	144.06	4.71	7797	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAT68762 +	137.50	138.65	9.41	9170
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA33184 +	140.50	143.80	4.86	7996	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA521391 +	137.50	134.39	16.25	1312
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA23895 +	140.50	125.12	53.35	49999	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC35941 +	137.00	157.85	0.8023	1310
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA23107 +	140.00	162.17	0.4610	1239	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAD02345 +	137.00	152.83	1.53	2144
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC76302 +	140.00	158.83	0.7072	1719	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA88863 +	137.00	150.55	2.05	2682
/SIDS1/gcgdata/geneseq/NA2001.DAT.AAC76212 +	140.00	150.49	2.06	3901	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC51132 +	137.00	145.83	3.75	4263
/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA54990 +	140.00	149.09	2.47	4475	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAC85085 +	137.00	145.35	3.99	4467
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA730092 +	140.00	146.98	3.23	5502	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAC32508 +	137.00	144.07	4.70	5068
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA831091 +	139.50	161.57	0.4979	1236	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA59124 +	137.00	142.67	5.62	5813
/SIDS1/gcgdata/geneseq/NA1995.DAT.AAA083526 +	139.50	155.30	1.11	2288	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA722450 +	137.00	142.65	5.63	5820
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA72894 +	139.50	155.30	1.11	2288	/SIDS1/gcgdata/geneseq/NA1992.DAT.AAO29187 +	136.50	151.84	1.73	2223
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA28283 +	139.50	155.30	1.11	2288	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA59640 +	136.50	150.44	2.08	2550
/SIDS1/gcgdata/geneseq/NA1993.DAT.AAA046543 +	139.50	154.88	1.17	2384	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV0553 +	136.50	146.38	3.49	3798
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV09624 +	139.50	154.88	1.17	2384	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT17473 +	136.50	141.88	6.22	5910
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA288975 +	139.50	146.56	3.41	5395	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT17474 +	136.50	141.88	6.22	5910
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT17450 +	139.50	145.62	3.85	5913	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT17475 +	136.50	141.88	6.22	5910
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA68403 +	139.50	143.09	5.35	7585	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA70206 +	136.50	139.44	8.50	7503
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA93712 +	139.50	127.82	37.78	39554	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC42748 +	136.50	137.95	10.30	8688
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA81484 +	139.50	116.55	160.30	102634	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA22301 +	136.50	136.99	11.66	9551
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA721609 +	139.50	104.05	796.25	349980	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA98272 +	136.50	136.67	12.13	9848
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA81490 +	139.50	89.66	5.0e+03	1437668	/SIDS1/gcgdata/geneseq/NA1991.DAT.AAQ13593 +	136.50	136.65	12.17	9871
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA72432 +	139.00	162.63	0.4349	1048	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA231401 +	136.00	157.33	0.8574	1219
/SIDS1/gcgdata/geneseq/NA1989.DAT.AAA94725 +	139.00	160.27	0.5879	1320	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA234451 +	136.00	157.33	0.8574	1219
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAAV83134 +	139.00	154.52	1.23	2322	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV64162 +	136.00	154.01	1.31	1689
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA61631 +	139.00	150.77	1.99	3354	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA40065 +	136.00	153.04	1.49	1857
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA093345 +	139.00	147.97	2.85	4416	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA234097 +	136.00	152.85	1.52	1834
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA093345 +	139.00	147.97	2.85	4416	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC78508 +	136.00	152.84	1.52	1894
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC75910 +	139.00	146.87	3.28	4920	/SIDS1/gcgdata/geneseq/NA2001.DAT.AAA32044 +	136.00	151.77	1.75	2105
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA79680 +	139.00	145.03	4.15	5992	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT13955 +	136.00	150.42	2.08	2403
/SIDS1/gcgdata/geneseq/NA1996.DAT.AAT17470 +	139.00	145.00	4.17	5913	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAT174004 +	136.00	150.42	2.08	2403
/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV27342 +	138.50	159.34	0.6624	1360	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV68065 +	136.00	150.42	2.08	2403
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA75781 +	138.50	152.84	1.53	2575	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV16177 +	136.00	150.42	2.08	2403
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC76844 +	138.50	151.81	1.74	2848	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA13964 +	136.00	150.25	2.13	2444
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC43193 +	138.50	148.79	2.56	3831	/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA74011 +	136.00	150.25	2.13	2444
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAC49498 +	138.50	148.33	2.72	4011	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAAV16190 +	136.00	150.25	2.13	2444
/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA403024 +	138.50	147.08	3.19	4534	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA84367 +	136.00	150.05	2.18	2492
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA093324 +	138.50	147.08	3.19	4534	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA90855 +	136.00	146.45	3.46	3546
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA7863781 +	138.50	123.97	61.87	43804	/SIDS1/gcgdata/geneseq/NA1999.DAT.AAA23429 +	136.00	146.41	3.48	3563
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA42007 +	138.00	165.57	0.2980	694	/SIDS1/gcgdata/geneseq/NA1998.DAT.AAA61588 +	136.00	146.31	3.52	3596
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA257866 +	138.00	158.49	0.7391	1391	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14682 +	136.00	144.22	4.61	4414
/SIDS1/gcgdata/geneseq/NA1991.DAT.AAA73115 +	138.00	150.66	2.02	3000	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14684 +	136.00	144.22	4.61	4414
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA73287 +	138.00	149.96	2.21	3214	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA14687 +	136.00	144.22	4.61	4414
/SIDS1/gcgdata/geneseq/NA1997.DAT.AAA73287 +	138.00	149.96	2.21	3214	/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA298401 +	136.00	143.48	5.07	4709
/SIDS1/gcgdata/geneseq/NA2000.DAT.AAA30290 +	138.00	149.12	2.46	3489	/SIDS1/gcgdata/geneseq/NA1996.DAT.AAA42028 +	136.00	141.60	6.45	5709











THIS PAGE BLANK (USPTO)



```

[1]
SEQUENCE FROM N.A.
RA STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RT Nature 402:761-768(1999).
DR EMBL; AC006201; AAD20129.1; -.
SQ SEQUENCE 507 AA; 57427 MW; B073165433A764CF CRC64;

Query Match 0.6%; Score 8; DB 10; Length 507;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332
Db 153 EEEEEEE 160

[1]
SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
```

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC 1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR EMBL; AE003560; AAF50628.1; -.
DR FlyBase; FBgn0035726; CG9953.
DR InterPro; IPR000073; -.
DR InterPro; IPR000379; -.
DR Pfam; PF00561; abhydrolase_1.
SQ SEQUENCE 508 AA; 56861 MW; 4AF644136F22F00B CRC64;

Query Match 0.6%; Score 8; DB 5; Length 508;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 KPXTNNVL 1131
Db 431 KPXTNNVL 438

[1]
SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 2.2 Mb region in 19q13.1 containing the RYR
gene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005239; AAC25503.1; -.
SQ SEQUENCE 510 AA; 57466 MW; CACE73EDC7290CE8 CRC64;

Query Match 0.6%; Score 8; DB 4; Length 510;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEDEEEE 1331
Db 355 PEDEEEE 362

Search completed: September 25, 2001, 10:07:44
Job time: 175 sec
```



Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1324 PEEEEEE 1331

DB 98 PEEEEEE 105

RESULT 45

ID O43818 PRELIMINARY; PRT; 475 AA.

AC O43818;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 16, Last annotation update)

DE U3 SNORNP ASSOCIATED 55 KDA PROTEIN.

GN U3-55K.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RE SEQUENCE FROM N.A.

RX MEDLINE=98078705; PubMed=9418896;

RA Pluk H., Soffner J., Luhrmann R., Van Venrooij W.J.;

RT "CDNA cloning and characterization of the human U3 small nucleolar

ribonucleoprotein complex-associated 55-kilodalton protein.";

RL Mol. Cell. Biol. 18:488-498(1998).

DR EMBL; AJ001340; CAA04687.1; -.

DR InterPro; IPR001680; -.

DR Pfam; PF00400; WD40; 6.

DR PRINTS; PR00320; GPROTEINBRPT.

DR PROSITE; PS00678; WD-REPEATS; UNKNOWN\_1.

DR SMART; SM00320; WD40; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 475 AA; 51840 MW; 3CD19F66EA75B627 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 4; Length 475;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1324 PEEEEEE 1331

DB 63 PEEEEEE 70

RESULT 46

ID Q39482 PRELIMINARY; PRT; 501 AA.

AC Q39482;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE LEGUMIN (FRAGMENT).

GN LEG1.

OS Calocedrus decurrens.

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Coniferopsida; Coniferales; Cupressaceae; Calocedrus.

OX NCBI\_TaxID=13387;

RN [1]

RE SEQUENCE FROM N.A.

RC TISSUE=IMMATURE SEED;

RA Haeger K.P., Dank N.;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; X95539; CAA64786.1; -.

DR Mendel; 7560; Calde; 1189; 7560.

DR InterPro; IPR000459; -.

DR InterPro; IPR001596; -.

DR Pfam; PF00190; Seedstore\_11s; 3.

DR PRINTS; PR00439; 11SGLOBULIN.

DR PROSITE; PS00387; PPASE; UNKNOWN\_1.

KW Seed storage protein.

FT NON\_TER 1

FT NON\_TER 1

SQ SEQUENCE 501 AA; 56975 MW; 7877D42C60E9D6E5 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 10; Length 501;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 EEEEEEE 1332

DB 140 EEEEEEE 147

RESULT 47

ID P97770 PRELIMINARY; PRT; 505 AA.

AC P97770;

DT 01-MAY-1997 (TREMBlrel. 03, Created)

DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)

DE HYPOTHETICAL 56.4 KDA PROTEIN (ROSA26AS).

GN GTROSA26AS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RE SEQUENCE FROM N.A.

RX MEDLINE=91357468; PubMed=1653172;

RA Friedrich G.A., Soriano P.;

RT "Promoter traps in embryonic stem cells: a genetic screen to identify

and mutate developmental genes in mice.";

RL Genes Dev. 5:1513-1523(1991).

RN [2]

RE SEQUENCE FROM N.A.

RX MEDLINE=97268648; PubMed=9108056;

RA Zambrowicz B.P., Inamoto A., Fiering S., Herzenberg L.A., Kerr W.G.,

Soriano P.;

RT "Disruption of overlapping transcripts in the ROSA beta geo 26 gene

trap strain leads to widespread expression of beta-galactosidase in

mouse embryos and hematopoietic cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:3789-3794(1997).

DR EMBL; U83176; AAC60384.1; -.

DR MGD; MGI:1277973; Gtrosa26as.

DR InterPro; IPR000241; -.

DR Pfam; PF01170; UPF0020; 1.

DR PROSITE; PS01261; UPF0020; UNKNOWN\_1.

KW Hypothetical protein.

SQ SEQUENCE 505 AA; 56431 MW; 6DA0F4C5B768C541 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 11; Length 505;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 221 LLKRTAQA 228

DB 483 LLKRTAQA 490

RESULT 48

ID Q9SL49 PRELIMINARY; PRT; 507 AA.

AC Q9SL49;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)

DE AT2G17970 PROTEIN.

GN AT2G17970.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;



DE CHROMOGRANIN A PRECURSOR.  
GN CHGA.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NEW ENGLAND DEACONNESS HOSPITAL;

RX MEDLINE=90007662; PubMed=2793216;

RA Farmer R.J., Koop A.H., Handa M.T., O'Connor D.T.;

RT "Molecular cloning of chromogranin A from rat pheochromocytoma cells";

RL Hypertension 14:435-444(1989).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NEW ENGLAND DEACONNESS HOSPITAL;

RA Farmer R.J., Koop A.H., Handa M.T., O'Connor D.T.;

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF145445; AAD40652.1; -

DR HSP; P05059; 1CFK

DR InterPro; IPR001819; -

DR Pfam; PF01271; Granin; 1.

DR PRINTS; PR00659; CHROMOGRANIN.

DR PROSITE; PS00422; GRANINS\_1; 1.

DR PROSITE; PS00423; GRANINS\_2; 1.

KW Signal.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 381 401 CATESTATIN.

SQ SEQUENCE 462 AA; 51512 MW; DA3F253DEC3F6282 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 11; Length 462;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332

DB 340 EEEEEEE 347

RESULT 42

Q9HEL6

ID Q9HEL6 PRELIMINARY; PRT; 466 AA.

AC Q9HEL6;

DT 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE CONSERVED HYPOTHETICAL PROTEIN.

GN 12F11.100.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL451017; CAC18219.1; -

SQ SEQUENCE 466 AA; 51449 MW; ED901EFCEC7ADCBD CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 3; Length 466;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1330

DB 448 APEEEEE 455

RESULT 43

OS7853

ID O57853 PRELIMINARY; PRT; 468 AA.

AC O57853;

DT 01-AUG-1998 (TRENBLrel. 07, Created)

DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE 468AA LONG HYPOTHETICAL REPLICATION FACTOR C SUBUNIT.

GN PH0113.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI\_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,

RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.;"

RL DNA Res. 5:55-76(1998).

DR EMBL; AF000001; BAA29182.1; -

DR InterPro; IPR000862; -

DR InterPro; IPR001939; -

DR InterPro; IPR003593; -

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 1.

SQ SEQUENCE 468 AA; 53913 MW; B85739AD2760B5AF CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 1; Length 468;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332

DB 436 EEEEEEE 443

RESULT 44

Q9QYK4

ID Q9QYK4 PRELIMINARY; PRT; 470 AA.

AC Q9QYK4;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE HEPARAN SULFATE 6-SULFOTRANSFERASE 3.

GN HS6ST3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Habuchi H.;

RT "Mouse, heparan sulfate 6-O-sulfotransferase.;"

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB024567; BAA89249.1; -

DR MGD; MGI:1354960; Hs6st3.

KW Transferase.

SQ SEQUENCE 470 AA; 55071 MW; 2D5EB0A82B895466 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 11; Length 470;

Matches 8; Conservative 100.0%; Pred. No. 40;



DR Pfam; PF00621; RhoGEF; 1.  
 FT NON\_TER 425 425  
 SQ SEQUENCE 425 AA; 49871 MW; C718616905736E7F CRC64;

Query Match 0.6%; Score 8; DB 4; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 DQ 11111111  
 57 PEEEEEE 64

RESULT 39  
 O34631 PRELIMINARY; PRT; 442 AA.

AC O34631;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE YVRA PROTEIN.  
 GN YVRA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Gulseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kashahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaite Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandembol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis.";  
 RT Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Wipat A., Brignell C.S., Guy J.B., Rose M., Emmerson P.T.,  
 RA Harwood C.R.;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Rose M., Entian K.;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z99120; CAB15306.1; -;  
 DR EMBL; Z99121; CAB15321.1; -;  
 DR EMBL; AJ223978; CAA11737.1; -;  
 DR InterPro; IPR001687; -;  
 DR InterPro; IPR003439; -;  
 DR InterPro; IPR003593; -;  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR SMART; SM00382; AAA; 1.  
 SQ SEQUENCE 442 AA; 48631 MW; 8020C6870E43C7AE CRC64;

Query Match 0.6%; Score 8; DB 2; Length 442;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 KDLLDLIK 514  
 DQ 11111111  
 174 KDLLDLIK 181

RESULT 40  
 Q9WVR7 PRELIMINARY; PRT; 450 AA.

AC Q9WVR7;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CA/CALMODULIN-DEPENDENT PROTEIN KINASE PHOSPHATASE.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99279559; PubMed=10348902;  
 RA Kitani T., Ishida A., Okuno S., Takeuchi M., Kameshita I.,  
 RA Fujisawa H.;  
 RT "Molecular cloning of Ca2+/Calmodulin-dependent protein kinase  
 RT phosphatase.";  
 RL J. Biochem. 125:1022-1028(1999).  
 DR EMBL; AB023634; BAA82477.1; -;  
 DR InterPro; IPR000222; -;  
 DR InterPro; IPR001932; -;  
 DR Pfam; PF00481; PP2C; 1.  
 DR PROSITE; PS01032; PP2C; 1.  
 DR SMART; SM00331; PP2C\_SIG; 1.  
 KW Kinase.  
 SQ SEQUENCE 450 AA; 49165 MW; 67626542F32B2FD0 CRC64;

Query Match 0.6%; Score 8; DB 11; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 DQ 11111111  
 103 EEEEEER 110

RESULT 41  
 Q9R1B7 PRELIMINARY; PRT; 462 AA.

AC Q9R1B7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)



KW Hypothetical protein.  
SQ SEQUENCE 379 AA; 40158 MW; CCE6BF58679A8142 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 379;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
DB 99 EEEEEER 106

RESULT 35  
Q9GL18 PRELIMINARY; PRT; 388 AA.  
AC Q9GL18;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE ALPHA ADRENERGIC RECEPTOR 2B (FRAGMENT).  
GN AAR2B.  
OS Dicerus bicornis (Black rhinoceros).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Dicerus.  
OX NCBI\_TaxID=9805;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Madsen O., Kao D., De Bry R.W., Douady C., Stanhope M.J.,  
de Jong W.W., Springer M.S.;  
RT "Molecules untangle the basal divergences of placental mammals.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ251184; CAC16686.1;  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 388 388  
SQ SEQUENCE 388 AA; 42546 MW; 14EB08C6FB4C5371 CRC64;

Query Match 0.6%; Score 8; DB 6; Length 388;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331  
DB 279 PEEEEER 286

RESULT 36  
Q9GK25 PRELIMINARY; PRT; 389 AA.  
AC Q9GK25;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE ALPHA ADRENERGIC RECEPTOR 2B (FRAGMENT).  
GN AAR2B.  
OS Tupia tana (large tree shrew).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Scandentia; Tupaiidae; Tupia.  
OX NCBI\_TaxID=70687;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Madsen O., Kao D., De Bry R.W., Douady C., Stanhope M.J.,  
de Jong W.W., Springer M.S.;  
RT "Molecules untangle the basal divergences of placental mammals.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ251187; CAC16698.1;  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 389 389  
SQ SEQUENCE 389 AA; 42631 MW; A9DC6F864487A2BE CRC64;

Query Match 0.6%; Score 8; DB 6; Length 389;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331  
DB 281 PEEEEER 288

RESULT 37  
Q9LTR6 PRELIMINARY; PRT; 395 AA.  
AC Q9LTR6;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE GB|AAF36750.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
clones.";  
RT DNA Res. 7:131-135(2000).  
DR EMBL; AB024036; BAB02807.1;  
SQ SEQUENCE 395 AA; 45347 MW; C4419188AD62622D CRC64;

Query Match 0.6%; Score 8; DB 10; Length 395;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331  
DB 79 PEEEEER 86

RESULT 38  
Q9NQD6 PRELIMINARY; PRT; 425 AA.  
AC Q9NQD6;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR (FRAGMENT).  
GN GEF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Rodrigues N.R., Theodosiou A.M., Nesbit M.A., Tandle A.T.,  
Saranath D., Davies K.E.;  
RT "A novel member of the Dbl family of oncogenes expressed predominantly  
in the caudate nucleus.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ238899; CAC00686.1;  
DR InterPro; IPR000219;



```

RESULT 31
Q9YFQ6 PRELIMINARY; PRT; 341 AA.
AC Q9YFQ6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 36.4 KDA PROTEIN APE0193.
GN APE0193.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RH SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310139; PubMed=10382966;
RA Kavarabavasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
RL EMBL; AP000058; BAA79105.1; -.
KW Hypothetical protein.
SQ SEQUENCE 341 AA; 36378 MW; ACD9D8D78A438B85 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 981 EKLISLLP 988
Db 325 EKLISLLP 332

RESULT 32
O96263 PRELIMINARY; PRT; 351 AA.
AC O96263;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RNA METHYLASE (SPOU FAMILY) (OO, TP).
GN PFE0855C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RH SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RL EMBL; AE001421; AAC71960.1; -.
DR InterPro; IPR001537; -.
DR Pfam; PF00588; Spou_methylase; 1.
DR ProDom; PD001243; -. 1.
DR Methyltransferase.
SQ SEQUENCE 351 AA; 41783 MW; 18EA09F50EFA4396 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 351;
Best Local Similarity 100.0%; Pred. No. 31;

```

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1270 LKEDILEN 1277
Db 305 LKEDILEN 312

RESULT 33
Q9P837 PRELIMINARY; PRT; 368 AA.
AC Q9P837;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACTIN BINDING PROTEIN (FRAGMENT).
GN ABP1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RH SEQUENCE FROM N.A.
RA De Backer M.D., Logghe M., Viaene J., Loonen I., Vandoninck S.,
RA de Hoogt R., Nelissen B., Dewaele S., Simons F., Verhasselt P.,
RA Contreras R., Luytens W.H.M.L.;
RT "A novel method for systematic identification of genes required for
RT growth of Candida albicans.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ390503; CAB77642.1; -.
DR InterPro; IPR000108; -.
DR InterPro; IPR001452; -.
DR InterPro; IPR002965; -.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 2.
FT NON_TER 1
SQ SEQUENCE 368 AA; 40241 MW; E1FC753CF1DC197F CRC64;

Query Match 0.6%; Score 8; DB 3; Length 368;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 176 PEEEEEE 183

RESULT 34
Q9XE25 PRELIMINARY; PRT; 379 AA.
AC Q9XE25;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 40.2 KDA PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RH SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
RT clone: P0026F07.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AP000364; BAA81779.1; -.
DR InterPro; IPR000847; -.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.

```



```

QY 1325 EEEEEER 1332
DB 114 EEEEEER 121

RESULT 28
O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE HYPOTHETICAL 38.4 KDA PROTEIN (G5 ORF).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OA NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RA "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
RA the Ddp1 and Ddp2 plasmid families.";
RA Genetics 148:1117-1125(1998).
RI EMBL; U00796; AAC18634.1; -
RW Hypothetical protein.
SQ SEQUENCE 325 AA; 38447 MW; 69A43D0C632058A6 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 325;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 159 PEEEEEE 166

RESULT 29
Q9FM77 PRELIMINARY; PRT; 329 AA.
AC Q9FM77;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MDF20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OA NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RA "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RA Sequence features of the regions of 1,456,315 bp covered by nineteen
RA physically assigned pl and TAC clones.";
RA DNA Res. 5:41-54(1998).
RI EMBL; AB09050; BAB09228.1; -
RW SEQUENCE 329 AA; 36055 MW; A5AA93B33FD7A44B CRC64;

Query Match 0.6%; Score 8; DB 10; Length 329;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 114 EEEEEER 121

```

```

DB 249 PEEEEEE 256

RESULT 30
Q9VD43 PRELIMINARY; PRT; 331 AA.
AC Q9VD43;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE CG7087 PROTEIN.
GN CG7087.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OA NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003738; AAF55960.1; -
DR FlyBase; FBgn0038935; CG7087.
SQ SEQUENCE 331 AA; 36524 MW; 3A805B2FA07468F6 CRC64;

```

```

Query Match 0.6%; Score 8; DB 5; Length 331;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 287 EEEEEER 294

```



GN PAC3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=WS;  
 RX MEDLINE=95003710; PubMed=7919990;  
 RA Reiter R.S., Comber S.A., Bourett T.M., Bartley G.E., Scolnik P.A.;  
 RT "Control of leaf and chloroplast development by the Arabidopsis gene  
 pale cress.";  
 RL Plant Cell 6:1253-1264(1994).  
 DR EMBL; L35241; AAA21761.1; -;  
 DR Mendel; 17382; Arath;2781;17382.  
 SQ SEQUENCE 310 AA; 36145 MW; 47567975BEFFFC41 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 DB 80 EEEEEER 87

RESULT 25  
 ID Q39089 PRELIMINARY; PRT; 313 AA.  
 AC Q39089;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE PALE CRESS GENE.  
 GN PALE CRESS OR AT2G48120.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=96323027; PubMed=8709959;  
 RA Greveling C., Suter-Crazzolara C., Menges A., von Kempner E.,  
 RA Masterson R., Schell J., Reiss B.;  
 RT "Characterisation of a new allele of pale cress and its role in  
 greening in Arabidopsis thaliana.";  
 RL Mol. Gen. Genet. 251:532-541(1996).  
 [2]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
 RL Nature 402:761-768(1999).  
 DR EMBL; X96480; CAA65334.1; -;  
 DR EMBL; X96482; CAA65337.1; -;  
 DR EMBL; AC006072; AAA48512.1; -;  
 DR Mendel; 17381; Arath;2781;17381.  
 SQ SEQUENCE 313 AA; 36611 MW; E3711AA2A4E975BE CRC64;

Query Match 0.6%; Score 8; DB 10; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332  
 DB 80 EEEEEER 87

RESULT 26  
 ID Q50887 PRELIMINARY; PRT; 316 AA.  
 AC Q50887;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CARD GENE AND OPEN READING FRAMES.  
 GN CARD.  
 OS Myxococcus xanthus.  
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
 OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.  
 OX NCBI\_TaxID=34;  
 [1]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=DK1050;  
 RX MEDLINE=96293442; PubMed=8692912;  
 RA Nicolas F.J., Cayuela M.L., Martinez-Argudo I.M., Ruiz-Vazquez R.M.,  
 RA Murillo F.J.;  
 RT "High mobility group I(Y)-like DNA-binding domains on a bacterial  
 transcription factor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:6881-6885(1996).  
 DR EMBL; 256280; CAA91224.1; -;  
 DR InterPro; IPR000637; -;  
 DR Pfam; PF02178; AT\_hook; 4.  
 DR PRINTS; PR00929; ATHOOK.  
 DR SMART; SM00384; AT\_hook; 1.  
 SQ SEQUENCE 316 AA; 34029 MW; B4277B36B2A60858 CRC64;

Query Match 0.6%; Score 8; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 KGRPPKP 1296  
 DB 257 KGRPPKP 264

RESULT 27  
 ID Q9LI19 PRELIMINARY; PRT; 316 AA.  
 AC Q9LI19;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
 OC Oryza.  
 OX NCBI\_TaxID=4530;  
 [1]  
 RE SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0708G02.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001539; BAA92919.1; -;  
 SQ SEQUENCE 316 AA; 33263 MW; 7D2BFA7A90C9C8B0 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-JUN-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE DOF ZINC FINGER PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
 OC Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YUKTHIKARI; TISSUE=ALEURONE LAYER;  
 RX MEDLINE=99418329; PubMed=10490396;  
 RA Washio K.;  
 RT "CDNAs encoding Dof-proteins that are present in germinated aleurone  
 cells (Accession Nos. AB028129, AB028130, AB028131, AB028132, and  
 AB028133). (PGR99-107).";  
 RL Plant Physiol. 120:1205-1205(1999).  
 DR EMBL; AB028133; BAA78576.1; -.  
 DR InterPro: IPR000345; -.  
 DR PROSITE: PS00190; CYNTOCHROME\_C; UNKNOWN1.  
 SQ SEQUENCE 235 AA; 25410 MW; D5B383072B30F540 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 235;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 DB 53 PEEEEEE 60

RESULT 21  
 O95472 PRELIMINARY; PRT; 251 AA.  
 ID O95472;  
 AC O95472;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE SLOW SKELETAL MUSCLE TROPONIN T.  
 GN TNNT1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Barton P.J.R., Cullen M.E., Townsend P.J., Brand N.J., Mullen A.J.,  
 RA Norman D.A.M., Bhavsar P.K., Yacoub M.H.;  
 RT "Close physical linkage of human troponin genes: organization,  
 RT sequence and expression of the locus encoding cardiac troponin I and  
 RT slow skeletal troponin T.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ011712; CAA09750.1; -.  
 DR EMBL; AJ011713; CAA09750.1; JOINED.  
 DR InterPro: IPR001978; -.  
 DR Pfam: PF00592; Troponin; 1.  
 SQ SEQUENCE 251 AA; 30096 MW; DA9E80D5F8B5DE19 CRC64;

Query Match 0.6%; Score 8; DB 4; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332  
 DB 21 EEEEEER 28  
 RESULT 22  
 O22924 PRELIMINARY; PRT; 294 AA.  
 ID O22924

AC O22924;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)  
 DE T09D09.9 PROTEIN.  
 GN T09D09.9  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC002338; AAC16933.1; -.  
 SQ SEQUENCE 294 AA; 34018 MW; 06217721EEEF2402 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 DB 226 PEEEEEE 233

RESULT 23  
 O9M4F1 PRELIMINARY; PRT; 299 AA.  
 ID O9M4F1;  
 AC O9M4F1;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE SIAH2 PROTEIN.  
 GN SIAH2.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. W1;  
 RA Cui Y., Brugiere N., Jackman L., Bi Y.M., Rothstein S.J.;  
 RT "Structural and transcriptional comparative analysis of the S-locus  
 RT regions in two self-incompatible Brassica napus lines.";  
 RL Plant Cell 11:2217-2231(2000).  
 DR EMBL; AJ245479; CAB89184.1; -.  
 SQ SEQUENCE 299 AA; 33686 MW; D5BEC8646E3111DD CRC64;

Query Match 0.6%; Score 8; DB 10; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331  
 DB 25 PEEEEEE 32

RESULT 24  
 O39175 PRELIMINARY; PRT; 310 AA.  
 ID O39175;  
 AC O39175;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE PALE CRESS PAC3.



RE SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RE SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RA PubMed=10907853;  
RN Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
TAC and BAC clones";  
RL DNA Res. 7:217-221(2000).  
DR EMBL; AB020746; BAB01999.1; -. 1A04100A17199268 CRC64;  
SQ SEQUENCE 149 AA; 17056 MW; 1A04100A17199268 CRC64;  
  
Query Match 0.6%; Score 8; DB 10; Length 149;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1324 PEEEEEE 1331  
DB 119 PEEEEEE 126  
  
RESULT 17  
Q23872 PRELIMINARY; PRT; 197 AA.  
ID Q23872;  
AC Q23872;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DR 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE G5/D6 ORF.  
OS Dictyostelium discoideum (Slime mold).  
OC Plasmid Dopl.  
CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RE SEQUENCE FROM N.A.  
RC STRAIN-NC4;  
RX MEDLINE=94302138; PubMed=8029325;  
RA Farrar N.A., Kiyosawa H., Hughes J.E., Welker D.L., Williams K.L.;  
RT "Nucleotide sequence of Dopl, a high copy number nuclear plasmid of  
Dictyostelium discoideum";  
RL Plasmid 31:184-195(1994).  
DR EMBL; U00691; AAC18627.1; -.  
KW Plasmid.  
SQ SEQUENCE 197 AA; 23224 MW; CD86BDEB9245CCB5 CRC64;  
  
Query Match 0.6%; Score 8; DB 5; Length 197;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1324 PEEEEEE 1331  
DB 49 PEEEEEE 56  
  
RESULT 18  
Q9SB54 PRELIMINARY; PRT; 204 AA.  
ID Q9SB54;  
AC Q9SB54;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DR 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 23.0 KDA PROTEIN.  
GN F22K18.210 OR AT4G24590.  
OS Arabidopsis thaliana (Mouse-ear cress).  
CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
BR Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;  
RN [1]  
RE SEQUENCE FROM N.A.  
RA Bevan M., Wedler H., Wedler E., Wambutt R., Hoheisel J., Mewes H.W.,  
RL Mayer K.F.X., Schueller C.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RE SEQUENCE FROM N.A.  
RN [3]  
RA EU Arabidopsis sequencing project;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RE SEQUENCE FROM N.A.  
RN [4]  
RA Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RE SEQUENCE FROM N.A.  
RN [5]  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL035356; CAA23004.1; -.  
DR EMBL; AL161561; CAB79369.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 204 AA; 22974 MW; D00A58B6EBA5C6FE CRC64;  
  
Query Match 0.6%; Score 8; DB 10; Length 204;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1324 PEEEEEE 1331  
DB 60 PEEEEEE 67  
  
RESULT 19  
Q9P529 PRELIMINARY; PRT; 231 AA.  
ID Q9P529;  
AC Q9P529;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DR 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 26.5 KDA PROTEIN.  
GN B24H17.160.  
OS Neurospora crassa.  
CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RE SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RL Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RE SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL356815; CAB92638.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 231 AA; 26523 MW; 9EA023CD8486583B CRC64;  
  
Query Match 0.6%; Score 8; DB 3; Length 231;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1325 EEEEEEE 1332  
DB 70 EEEEEEE 77  
  
RESULT 20  
Q9SXG4 PRELIMINARY; PRT; 235 AA.  
ID Q9SXG4  
AC Q9SXG4;



```
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 99 AA; 10472 MW; 5BA44A226B8A6990 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 99;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 KGRGPRSK 1380
DB 55 KGRGPRSK 62

RESULT 13
Q23794 PRELIMINARY; PRT; 99 AA.
ID Q23794
AC Q23794;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN I/Y.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95105193; PubMed=7806532;
RA Claus P., Schulze E., Wisniewski J.R.;
RT "Insect proteins homologous to mammalian high mobility group proteins
RT I/Y (HMG I/Y). Characterization and binding to linear and four-way
RT junction DNA."; 269:33042-33048(1994).
RL J. Biol. Chem. 269:33042-33048(1994).
DR EMBL; Z36898; CAA85365.1; -
DR InterPro; IPR000637; -
DR Pfam; PF02178; AT_hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 99 AA; 10502 MW; 1AA44A226B8A698C CRC64;

Query Match 0.6%; Score 8; DB 5; Length 99;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 KGRGPRSK 1380
DB 55 KGRGPRSK 62

RESULT 14
Q96475 PRELIMINARY; PRT; 131 AA.
ID Q96475
AC Q96475;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CONGLUTIN ALPHA (CONALPHA) (FRAGMENT).
GN CONALPHA.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Lupinus.
OX NCBI_TaxID=3871;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UNICROP.
RA MEDLINE=97390686; PubMed=9247543;
RA Ilgoutz S.C., Knittel N., Lin J.M., Sterle S., Gayler K.R.;
RT "Transcription of genes for conglutin gamma and a leginsulin-like
RT protein in narrow-leaved lupin.";

Plant Mol. Biol. 34:613-627(1997).
-!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER L1S SEED STORAGE PROTEINS (GLOBULINS).
DR EMBL; U74384; AAC49787.1; -
DR Mendel; 24412; Lupan; 1189; 24412.
DR InterPro; IPR000459; -
DR Pfam; PF00190; Seedstore_l1s; 1.
DR PRINTS; PR00439; L1SGLOBULIN.
DR PROSITE; PS00305; L1S_SEED_STORAGE; 1.
KW Multigene family.
FT NON_TER 1
FT NON_TER 131
SQ SEQUENCE 131 AA; 15532 MW; 75BF5C9609AD9F09 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 131;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 8 EEEEEER 15

RESULT 15
Q84290 PRELIMINARY; PRT; 147 AA.
ID Q84290
AC Q84290;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TYPE 4 E4 PROTEIN.
GN E4.
OS Human papillomavirus.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10566;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89370332; PubMed=2549722;
RA Doorbar J., Coneron I., Gallimore P.H.;
RT "Sequence divergence yet conserved physical characteristics among the
RT E4 proteins of cutaneous human papillomaviruses.";
RL Virology 172:51-62(1989).
DR EMBL; M28744; AAA47015.1; -
SQ SEQUENCE 147 AA; 16854 MW; A238A5DCEEC8C9FD CRC64;

Query Match 0.6%; Score 8; DB 14; Length 147;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 102 EEEEEER 109

RESULT 16
Q9LV56 PRELIMINARY; PRT; 149 AA.
ID Q9LV56
AC Q9LV56;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE GB|AAD25142.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
```



```
DR InterPro; IPR000637; -.
DR Pfam; PF02178; AT_hook; 3.
DR PRINTS; PRO0929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
DR SC SEQUENCE 837 AA; 90444 MW; 140C484C027036BF CRC64;

Query Match          0.6%; Score 9; DB 3; Length 837;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1287 KGKGRPPK 1295
Db          |||||
695 KGKGRPPK 703

RESULT 9
QSRIN5 PRELIMINARY; PRT; 59 AA.
AC QSRINS;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL; AF066061; AAD43365.1; -.
FT NON_TER 1
FT NON_TER 59
FT SEQUENCE 59 AA; 6878 MW; C64B99BD0D94826 CRC64;

Query Match          0.6%; Score 8; DB 11; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db          |||||
39 PEEEEEE 46

RESULT 10
QSWVP2 PRELIMINARY; PRT; 66 AA.
AC QSWVP2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL; AF066061; AAD43364.1; -.
FT NON_TER 1
FT NON_TER 66
FT SEQUENCE 66 AA; 7645 MW; 3B655092C1355B3B CRC64;

InterPro; IPR000637; -.
Pfam; PF02178; AT_hook; 3.
PRINTS; PRO0929; ATHOOK.
SMART; SM00384; AT_hook; 1.
SEQUENCE 837 AA; 90444 MW; 140C484C027036BF CRC64;

Query Match          0.6%; Score 8; DB 11; Length 66;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db          |||||
39 PEEEEEE 46

RESULT 11
Q9M3G5 PRELIMINARY; PRT; 81 AA.
AC Q9M3G5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 8.9 KDA PROTEIN.
GN T14K23_20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nyakatura G., Fartmann B., Dauner D., Sterr W., Holland R.,
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RA Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132909; CAB87723.1; -.
KW Hypothetical protein.
SQ SEQUENCE 81 AA; 8929 MW; 1B2371B2F081ABEE CRC64;

Query Match          0.6%; Score 8; DB 10; Length 81;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 PEEEEEE 1332
Db          |||||
19 PEEEEEE 26

RESULT 12
Q23793 PRELIMINARY; PRT; 99 AA.
AC Q23793;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN I/Y.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95105193; PubMed=7806532;
RA Claus P., Schulze E., Wisniewski J.R.;
RT "Insect proteins homologous to mammalian high mobility group proteins
RT I/Y (HMG I/Y): Characterization and binding to linear and four-way
RT junction DNA.";
RL J. Biol. Chem. 269:33042-33048(1994).
DR EMBL; Z36897; CAA85364.1; -.
FT InterPro; IPR000637; -.
DR Pfam; PF02178; AT_hook; 3.
```



Matches	9;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1323	APEEEEEE	1331						
DB	53	APEEEEEE	61						
RESULT	7								
Q9SG14									
ID	Q9SG14	PRELIMINARY;	PRT;	554	AA.				
AC	Q9SG14;								
DT	01-MAY-2000	(TEMBLrel. 13, Created)							
DT	01-MAY-2000	(TEMBLrel. 13, Last sequence update)							
DT	01-MAR-2001	(TEMBLrel. 16, Last annotation update)							
DE	PUTATIVE GTPASE ACTIVATOR PROTEIN OF RAB-LIKE SMALL GTPASES.								
GN	TIG12.5 OR F2K15.210.								
OS	Arabidopsis thaliana (Mouse-ear cross).								
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;								
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;								
OC	Brassicales; Brassicaceae; Arabidopsids.								
OX	NCBI_TaxID=3702;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=CV. COLOMBIA;								
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,								
RA	Maiti R., Renning C.M., Koo H., Fujii C.Y., Utterback T.R.,								
RA	Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;								
RT	"Arabidopsis thaliana chromosome 1 BAC TIG12 genomic sequence.";								
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,								
RA	Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salancoubat M.;								
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; AC013329; AAC52193.1; -								
DR	EMBL; AL132956; CAB66414.1; -								
DR	InterPro; IPR000195; -								
DR	Pfam; PF00366; TBC; 1.								
SQ	SEQUENCE 554 AA; 63583 MW; 430BE1CFID8901C CRC64;								
Query Match	0.6%;	Score 9;	DB 10;	Length 554;					
Best Local Similarity	100.0%;	Pred. No. 4.1;							
Matches	9;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1159	SSSSNPSSP	1167						
DB	11	SSSSNPSSP	19						
RESULT	8								
Q9P607									
ID	Q9P607	PRELIMINARY;	PRT;	837	AA.				
AC	Q9P607;								
DT	01-OCT-2000	(TEMBLrel. 15, Created)							
DT	01-OCT-2000	(TEMBLrel. 15, Last sequence update)							
DT	01-MAR-2001	(TEMBLrel. 16, Last annotation update)							
DE	NEUROFILAMENT-H RELATED PROTEIN.								
GN	B208.100.								
OS	Eukaryota; Crassa.								
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;								
OC	Sordariales; Sordariaceae; Neurospora.								
OX	NCBI_TaxID=5141;								



Proc. Am. Assoc. Cancer Res. 37:223-223(1996).

[12]

SEQUENCE FROM N.A.

TISSUE=PROSTATE;

MEDLINE=99229875; PubMed=10215036;

Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;

"Early gene expression during androgen-induced inhibition of

proliferation of prostate cancer cells: a new suppressor candidate on

chromosome 13, in the BRCA2-Rb1 locus.";

J. Steroid Biochem. Mol. Biol. 68:41-45(1999).

EMBL: U95825; AAD22134.2; -

InterPro: IPR002687; -

ProDom: PD004104; -

SEQUENCE 1391 AA; 158035 MW; F58ABEE5AD69479 CRC64;

Query Match 25.4%; Score 353; DB 4; Length 1391;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 853; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

257 DLILELYNTDLSLLSVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQ 316

|||||

257 DLILELYNTDLSLLSVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQ 316

|||||

317 CYLGRFNDIHPVIRLECYKFASHCLMNHDPDLAKDTEYLKVRSHDPEAIRHDVIVSYT 376

|||||

317 CYLGRFNDIHPVIRLECYKFASHCLMNHDPDLAKDTEYLKVRSHDPEAIRHDVIVSYT 376

|||||

377 AAKDILLVNDHLLNFVRERTLDKRRVRKEAMGLAQIYKYLQSAAGKDAAKQIAWI 436

|||||

377 AAKDILLVNDHLLNFVRERTLDKRRVRKEAMGLAQIYKYLQSAAGKDAAKQIAWI 436

|||||

437 KDKLLHIYQNSIDRLLVERIFAQYMPVPHNLETTERMKCLYLYATLDLNAVKAALNEMW 496

|||||

437 KDKLLHIYQNSIDRLLVERIFAQYMPVPHNLETTERMKCLYLYATLDLNAVKAALNEMW 496

|||||

497 KCONLLRHQVKDLDLILQPKTDASVKAIFSKVMVITRNLPDGGRAQDPKMKFTQVLEDD 556

|||||

497 KCONLLRHQVKDLDLILQPKTDASVKAIFSKVMVITRNLPDGGRAQDPKMKFTQVLEDD 556

|||||

557 EKIRKQELVLSPTSCQAQECVREITKGLGNPKQPTNPFLEIMKFLERLAPVHIDTE 616

|||||

557 EKIRKQELVLSPTSCQAQECVREITKGLGNPKQPTNPFLEIMKFLERLAPVHIDTE 616

|||||

617 SISALIKQVNSIDGTADDEGVPDQAIRAGLELLKVLSTHPTSFHSAETFSLLAC 676

|||||

617 SISALIKQVNSIDGTADDEGVPDQAIRAGLELLKVLSTHPTSFHSAETFSLLAC 676

|||||

677 LKMDDEKVAEAAALQIFKNTGSKIEEDFPHIRSAALLPVLLHKKGGPPROAKYAIHCHIAI 736

|||||

677 LKMDDEKVAEAAALQIFKNTGSKIEEDFPHIRSAALLPVLLHKKGGPPROAKYAIHCHIAI 736

|||||

737 FSSKETQFAQIEPELHKSIDPSNLEHLITPLVTIGHIALLAPDQFAAPKMSWATFIVKD 796

|||||

737 FSSKETQFAQIEPELHKSIDPSNLEHLITPLVTIGHIALLAPDQFAAPKMSWATFIVKD 796

|||||

797 LLMNDRLPGKTKTLWVPDEEVSPETWVKIQAKMVRWLLGMKNHNSKSGTSLRLTT 856

|||||

797 LLMNDRLPGKTKTLWVPDEEVSPETWVKIQAKMVRWLLGMKNHNSKSGTSLRLTT 856

|||||

857 ILHSDGDLTEQCKISKPDMSRLRGAAGSVAIVKLAQPCVHEITILEQVLCALAINDECY 916

|||||

857 ILHSDGDLTEQCKISKPDMSRLRGAAGSVAIVKLAQPCVHEITILEQVLCALAINDECY 916

|||||

917 QVRQVFAQKLHGLSLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQH 976

|||||

917 QVRQVFAQKLHGLSLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQH 976

|||||

977 AAVSEKLSLLPEYVVPYTHLLAHDPPYVKVDIEQLKDVKECLWFVLEILMAKNENNS 1036

|||||

977 AAVSEKLSLLPEYVVPYTHLLAHDPPYVKVDIEQLKDVKECLWFVLEILMAKNENNS 1036

|||||

1037 HAFIRKMWENIKQTKDAQGPDDAKMKNEKLYTVCVAMNIIIMSKSTYSLESKDPVLPAR 1096

Db 1037 HAFIRKMWENIKQTKDAQGPDDAKMKNEKLYTVCVAMNIIIMSKSTYSLESKDPVLPAR 1096

Qy 1097 FFTQDPKFNSTKNYLPP 1114

Db 1097 FFTQDPKFNSTKNYLPP 1114

RESULT 4

Q9H5N8

ID Q9H5N8 PRELIMINARY; PRT; 125 AA.

AC Q9H5N8

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DE CDNA: FLJ23236 FIS; CLONE COL00725.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=COLON;

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK026889; BAB15584.1; -

SQ SEQUENCE 125 AA; 14281 MW; 492809836D3665D3 CRC64;

Query Match 9.0%; Score 125; DB 4; Length 125;  
Best Local Similarity 100.0%; Pred. No. 1e-122;  
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1231 MDDLTKLVQEQPKGSQSRKRGHTASESDEQOWPEEKRLKEDILENEDEQNSPPKKGR 1290

Db 1 MDDLTKLVQEQPKGSQSRKRGHTASESDEQOWPEEKRLKEDILENEDEQNSPPKKGR 60

Qy 1291 GRPPPLGGGTPEKPTMTKTSKKGSKKSGPPAPPEEEEEESQNTQKSKQHRVSR 1350

Db 61 GRPPPLGGGTPEKPTMTKTSKKGSKKSGPPAPPEEEEEESQNTQKSKQHRVSR 120

Qy 1351 RAQQR 1355

Db 121 RAQQR 125

RESULT 5

Q9Y4D4

ID Q9Y4D4 PRELIMINARY; PRT; 851 AA.

AC Q9Y4D4;

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DE KIAA0648 PROTEIN (FRAGMENT).

GN KIAA0648.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRIN;

RX MEDLINE=98403880; PubMed=9734811;

RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,

RA Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. X.

RT The complete sequences of 100 new cDNA clones from brain which can

code for large proteins in vitro.";

RL DNA Res. 5:169-176(1998).

DR EMBL: AB014548; BAA31623.1; -



QX 1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGSQSRKRKRGHTASESDEQOWPEEKRLKED 1273  
 DB 1183 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGSQSRKRKRGHTASESDEQOWPEEKRLKED 1242  
 QX 1274 ILENEDEQNSPPKGGKRGPPKPLGGGTTPKEPTMTSKSGSKSGPPAPPEEEEEERQ 1333  
 DB 1243 ILENEDEQNSPPKGGKRGPPKPLGGGTTPKEPTMTSKSGSKSGPPAPPEEEEEERQ 1302  
 QX 1334 SGNTEOKSKSKQHRVSRRAQRAESPESSAIESTQSTPQKGRGRPKSTPSPQPKNV 1391  
 DB 1303 SGNTEOKSKSKQHRVSRRAQRAESPESSAIESTQSTPQKGRGRPKSTPSPQPKNV 1360  
 RESULT 2  
 Q9NT15 PRELIMINARY; PRT: 1447 AA.  
 AC Q9NT15;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DB HYPOTHETICAL 164.7 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rhodes S., Huckle E.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL137201; CAB69911.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1447 AA; 164666 MW; 145C30308EA3EFD5 CRC64;  
 Query Match - 73.0%; Score 1015; DB 4; Length 1447;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1315; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QX 74 KDVRLLVACCLADIFRIYAPAYTSPDKLKIDFMFTRQLKGLDTPKSPQNRNRYLYLE 133  
 DB 74 KDVRLLVACCLADIFRIYAPAYTSPDKLKIDFMFTRQLKGLDTPKSPQNRNRYLYLE 133  
 QX 134 NIAWKSYNICFELEDSNEIFTOLYTLFVINGNHQVHMVDMSSICEGDTVSQ 193  
 DB 134 NIAWKSYNICFELEDSNEIFTOLYTLFVINGNHQVHMVDMSSICEGDTVSQ 193  
 QX 194 ELLDTVLNVLPAHKNLKOAYDLAKALKRTAQAIPEYITTFNQVLMGLKTSISDLSE 253  
 DB 194 ELLDTVLNVLPAHKNLKOAYDLAKALKRTAQAIPEYITTFNQVLMGLKTSISDLSE 253  
 QX 254 HVPDLILELYNDSHLLSVLPOLKPKLSNDNEERLQVVKLLAKMFGAKDSLASQNK 313  
 DB 254 HVPDLILELYNDSHLLSVLPOLKPKLSNDNEERLQVVKLLAKMFGAKDSLASQNK 313  
 QX 314 LMOCYLGRENTHVPLRLECVKFASHCLMNHDPDLAKDLTKYLVKVRSHDPEAIRHDI 373  
 DB 314 LMOCYLGRENTHVPLRLECVKFASHCLMNHDPDLAKDLTKYLVKVRSHDPEAIRHDI 373  
 QX 374 IVTAARKDILLVNDHLLNFVRETLDRWRVRKEAMGLAQIYKYLQASAGKDAKQI 433  
 DB 374 IVTAARKDILLVNDHLLNFVRETLDRWRVRKEAMGLAQIYKYLQASAGKDAKQI 433  
 QX 434 AWTKDILLHIYQNSIDRLLRIFERFAQYVPHNLETTERMKCLLYLYATLDLNAVKN 493  
 DB 434 AWTKDILLHIYQNSIDRLLRIFERFAQYVPHNLETTERMKCLLYLYATLDLNAVKN 493  
 QX 494 EMWKNLRLHVOVKDILLIKPKTDASVKAIFSKVWVITRNLPDPGKADFMKKTQVL 553  
 DB 494 EMWKNLRLHVOVKDILLIKPKTDASVKAIFSKVWVITRNLPDPGKADFMKKTQVL 553  
 QX 554 EDEKIRKQLEVLVSPSCSKQAEQGVREITTKLGNPKQPTNPFLEMIKFLLERIAPVHI 613  
 DB 554 EDEKIRKQLEVLVSPSCSKQAEQGVREITTKLGNPKQPTNPFLEMIKFLLERIAPVHI 613

DB 554 EDEKIRKQLEVLVSPSCSKQAEQGVREITTKLGNPKQPTNPFLEMIKFLLERIAPVHI 613  
 QX 614 DTESISALIKQVNSKSDGTADDEDEGVPTDOATRAGLEILLKLVLSFTPTSFHSAETFESL 673  
 DB 614 DTESISALIKQVNSKSDGTADDEDEGVPTDOATRAGLEILLKLVLSFTPTSFHSAETFESL 673  
 QX 674 LACLKMDDEKVAEAAALQIFKNTGSKIETEDFPPIRSALLPVLHHKSKGPPROAKYAIHCI 733  
 DB 674 LACLKMDDEKVAEAAALQIFKNTGSKIETEDFPPIRSALLPVLHHKSKGPPROAKYAIHCI 733  
 QX 734 HAIFSSKETQFAQIFEPRLHKSIDPSNLEHLITPLVTIGHIALAPDOFAAPKWSWATFI 793  
 DB 734 HAIFSSKETQFAQIFEPRLHKSIDPSNLEHLITPLVTIGHIALAPDOFAAPKWSWATFI 793  
 QX 794 VKDLLMNDRLPGKKTTLKLVVDPDEEVSPTWVKIOAKIMVWRLGKNNHNSKSGTSLRL 853  
 DB 794 VKDLLMNDRLPGKKTTLKLVVDPDEEVSPTWVKIOAKIMVWRLGKNNHNSKSGTSLRL 853  
 QX 854 LTTILHSDGDLTEOGKISKPDMSRLRLAAGSAIVKLAQPCYHEIITLEQYQCALAIND 913  
 DB 854 LTTILHSDGDLTEOGKISKPDMSRLRLAAGSAIVKLAQPCYHEIITLEQYQCALAIND 913  
 QX 914 ECVQVQVFAQKLHKLGLSRLRPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYL 973  
 DB 914 ECVQVQVFAQKLHKLGLSRLRPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYL 973  
 QX 974 KQHAAYSEKLLSLLPEYVVPYTHLLAHPDYPVKVDIEQLKDVKECLFVLEILMAKNE 1033  
 DB 974 KQHAAYSEKLLSLLPEYVVPYTHLLAHPDYPVKVDIEQLKDVKECLFVLEILMAKNE 1033  
 QX 1034 NNSHAFIRKMWENIKOTKDAQGPDADKMEKLYTCDVAMNIIKMSKSTYTSLESKDPVL 1093  
 DB 1034 NNSHAFIRKMWENIKOTKDAQGPDADKMEKLYTCDVAMNIIKMSKSTYTSLESKDPVL 1093  
 QX 1094 PARFFTQDPKNFNTKNYLPPEMKSEFTPGKPTTNVLGAVNKPLSSACKQSQTSSRME 1153  
 DB 1094 PARFFTQDPKNFNTKNYLPPEMKSEFTPGKPTTNVLGAVNKPLSSACKQSQTSSRME 1153  
 QX 1154 TVSNASSSNPSPGRIKGLSDSEMDHSENYDTWSSPLPGKSKDKRSDSLVSRSELEK 1213  
 DB 1154 TVSNASSSNPSPGRIKGLSDSEMDHSENYDTWSSPLPGKSKDKRSDSLVSRSELEK 1213  
 QX 1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGSQSRKRKRGHTASESDEQOWPEEKRLKED 1273  
 DB 1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGSQSRKRKRGHTASESDEQOWPEEKRLKED 1273  
 QX 1274 ILENEDEQNSPPKGGKRGPPKPLGGGTTPKEPTMTSKSGSKSGPPAPPEEEEEERQ 1333  
 DB 1274 ILENEDEQNSPPKGGKRGPPKPLGGGTTPKEPTMTSKSGSKSGPPAPPEEEEEERQ 1333  
 QX 1334 SGNTEOKSKSKQHRVSRRAQRAESPESSAIESTQSTPQKGRGRPKSTPSPQPKNV 1391  
 DB 1334 SGNTEOKSKSKQHRVSRRAQRAESPESSAIESTQSTPQKGRGRPKSTPSPQPKNV 1391

RESULT 3  
 Q9Y451 PRELIMINARY; PRT: 1391 AA.  
 AC Q9Y451;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-NOV-2001 (Tremblrel. 16, Last annotation update)  
 DE ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN.  
 GN AS3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=PROSTATE;  
 RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;  
 RT "Androgen-induced proliferative shutoff in prostate cancer cells."



969 7 0.5 719 11 Q61281 mus musculus  
970 7 0.5 720 3 O14367 schizosacch  
971 7 0.5 721 5 Q9NAB9 caenorhabdi  
972 7 0.5 722 5 Q9VFK4 drosophila  
973 7 0.5 725 4 Q9UFA6 homo sapien  
974 7 0.5 725 4 Q9H986 homo sapien  
975 7 0.5 726 10 Q9FGN7 arabidopsis  
976 7 0.5 726 11 Q9QZV4 mus musculus  
977 7 0.5 730 14 Q06634 bovine herp  
978 7 0.5 732 14 Q9DUC7 Q9duc7 tt virus, o  
979 7 0.5 735 13 Q9DDQ9 Q9ddo9 xenopus lae  
980 7 0.5 736 6 O18805 o18805 cercopithe  
981 7 0.5 736 11 Q9ESN0 Q9esn0 rattus norv  
982 7 0.5 737 4 Q9UP58 Q9up58 homo sapien  
983 7 0.5 738 4 Q9UNY2 Q9uny2 homo sapien  
984 7 0.5 738 11 Q9WVE4 Q9wve4 mus musculus  
985 7 0.5 738 11 Q9QY02 Q9qy02 rattus norv  
986 7 0.5 739 10 Q9FTF9 Q9ftf9 arabidopsis  
987 7 0.5 739 11 Q35613 Q35613 mus musculus  
988 7 0.5 740 4 O15141 O15141 homo sapien  
989 7 0.5 740 4 Q9UER7 Q9uer7 homo sapien  
990 7 0.5 740 4 Q9UWJ0 Q9uwj0 homo sapien  
991 7 0.5 740 4 O14747 O14747 homo sapien  
992 7 0.5 740 11 Q9QWT8 Q9qwt8 mus musculus  
993 7 0.5 741 3 Q9Y7K6 Q9y7k6 schizosacch  
994 7 0.5 742 5 Q9NEB3 Q9neb3 leishmania  
995 7 0.5 742 10 Q9FW12 Q9fw12 oryza sativ  
996 7 0.5 748 4 Q9Y6E1 Q9y6e1 homo sapien  
997 7 0.5 748 4 Q9UBJ0 Q9ubj0 homo sapien  
998 7 0.5 750 14 Q9Y039 Q9yq39 pseudorabie  
999 7 0.5 751 5 Q9I7S7 Q9i7s7 drosophila  
1000 7 0.5 751 11 Q60709 Q60709 mus musculus

## ALIGNMENTS

## RESULT 1

Q9Y215 PRELIMINARY; PRT; 1416 AA.  
AC Q9Y215;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE KIAA0979 PROTEIN (FRAGMENT).  
GN KIAA0979.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RE SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99246063; PubMed=10231032;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:63-70(1999).  
DR EMBL; AB023196; BAA76823.1; -;  
DR InterPro; IPR000637; -;  
DR PRINTS; PR00929; ATHOOK.  
FT NON\_TER 1  
SQ SEQUENCE 1416 AA; 161192 MW; 0BC682D7F4657424 CRC64;

Query Match 73.0%; Score 1015; DB 4; Length 1416;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1315; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

74 KDVRLLVACCLADIPRIYAPEAPYTSPOKLDIFMFITROLKGLDTSKSPQNRVYFYLLE 133  
|||||

Db 43 KDVRLLVACCLADIPRIYAPEAPYTSPOKLDIFMFITROLKGLDTSKSPQNRVYFYLLE 102  
Qy 134 NIAWKSYNICFEDESDSNEIFTQLYRTLFSVINNGHGNOKVHMVMDLMSIIICEGDTVSQ 193  
Db 103 NIAWKSYNICFEDESDSNEIFTQLYRTLFSVINNGHGNOKVHMVMDLMSIIICEGDTVSQ 162  
Qy 194 ELLDVLVNLVPAHKNLKNQAYDLAKALLKRTAQAIIEPYITTFNVOVLMGLKTSISDLSE 253  
Db 163 ELLDVLVNLVPAHKNLKNQAYDLAKALLKRTAQAIIEPYITTFNVOVLMGLKTSISDLSE 222  
Qy 254 HVFDLILELYNIDSHLLSVLPQLEFKLSNDNERLQVVKLLAKMFGAKDSELAISONKP 313  
Db 223 HVFDLILELYNIDSHLLSVLPQLEFKLSNDNERLQVVKLLAKMFGAKDSELAISONKP 282  
Qy 314 LMQCYLGRFNIDHVPRIECLVKFASHCLMNHDPDLAKDTEYLKVRSHDPEEAIIRDVIYS 373  
Db 283 LMQCYLGRFNIDHVPRIECLVKFASHCLMNHDPDLAKDTEYLKVRSHDPEEAIIRDVIYS 342  
Qy 374 IVTAAKDDILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYAQSAAAGKDAAKOI 433  
Db 343 IVTAAKDDILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYAQSAAAGKDAAKOI 402  
Qy 434 AWIKDKLLHIYQNSIDDRLLVERIFAQVMVPHNLLETTERMKCLYLYATLDLNAVKAALN 493  
Db 403 AWIKDKLLHIYQNSIDDRLLVERIFAQVMVPHNLLETTERMKCLYLYATLDLNAVKAALN 462  
Qy 494 EMKQCNLLRHQVKDILLIKOPKTDASVKAIFSVMVITRNLDPGKAQDPKMKFTQVL 553  
Db 463 EMKQCNLLRHQVKDILLIKOPKTDASVKAIFSVMVITRNLDPGKAQDPKMKFTQVL 522  
Qy 554 EDEKIRKQLEVLVSPTCSQAQECVREITKKNLPQTPNPFLKFLERLAPVHI 613  
Db 523 EDEKIRKQLEVLVSPTCSQAQECVREITKKNLPQTPNPFLKFLERLAPVHI 582  
Qy 614 DTESISALIKOVNSIDGTADDEGVPTDQAIIRAGLELLKVLSTHPTSFHSAETFESL 673  
Db 583 DTESISALIKOVNSIDGTADDEGVPTDQAIIRAGLELLKVLSTHPTSFHSAETFESL 642  
Qy 674 LACLKMDDEKVAEAAALQIFKNTGSKIIEEDFPHIRISALLPVLHKKSGKGPPOAKYAIHCI 733  
Db 643 LACLKMDDEKVAEAAALQIFKNTGSKIIEEDFPHIRISALLPVLHKKSGKGPPOAKYAIHCI 702  
Qy 734 HAIFSSKETQFAQIEPELHKSLDPSNLEHLITPLVTIGHIALLAPDOFAAPKSWVATPI 793  
Db 703 HAIFSSKETQFAQIEPELHKSLDPSNLEHLITPLVTIGHIALLAPDOFAAPKSWVATPI 762  
Qy 794 VKDLMLNDRLPCKTKTKLWVPDEEVSPTMWKIOAKMMVRWLLGMKNHNSKSGTSTLRL 853  
Db 763 VKDLMLNDRLPCKTKTKLWVPDEEVSPTMWKIOAKMMVRWLLGMKNHNSKSGTSTLRL 822  
Qy 854 LTTILHSDGDLTEQCKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLEQYQCALAIND 913  
Db 823 LTTILHSDGDLTEQCKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLEQYQCALAIND 882  
Qy 914 ECVQYRVQFAQKLHKLRLPLEYMAICALKADPKVKEKRAHARQCLVKNNVREYL 973  
Db 883 ECVQYRVQFAQKLHKLRLPLEYMAICALKADPKVKEKRAHARQCLVKNNVREYL 942  
Qy 974 KOHAAVSEKLLSLLEPYVVPYTIHLLAHDPPYVKVDTEOLKDKVECLWFLVLEILMAKNE 1033  
Db 943 KOHAAVSEKLLSLLEPYVVPYTIHLLAHDPPYVKVDTEOLKDKVECLWFLVLEILMAKNE 1002  
Qy 1034 NNSHAFIRKMWENIKOTKDAQGDDAKMNEKLYTVCDVAMNIIIMSKSTTYSLESKPDVYL 1093  
Db 1003 NNSHAFIRKMWENIKOTKDAQGDDAKMNEKLYTVCDVAMNIIIMSKSTTYSLESKPDVYL 1062  
Qy 1094 PARFTQDPKNFNTKNYLPPEMKSFPTPGPKPTNVLGAVNKPPLSSAGKQSQTKSSRME 1153  
Db 1063 PARFTQDPKNFNTKNYLPPEMKSFPTPGPKPTNVLGAVNKPPLSSAGKQSQTKSSRME 1122  
Qy 1154 TVSNASSSSNPPSGRIKGRDLSSEMDHSENEDEYTWSSPLPGCKSKDKRDDSDLVVRSELEK 1213  
Db 1123 TVSNASSSSNPPSGRIKGRDLSSEMDHSENEDEYTWSSPLPGCKSKDKRDDSDLVVRSELEK 1182



823	7	0.5	580	10	Q92WB0	Q92wb0 arabidopsis	896	7	0.5	634	5	Q06279	Q06279 schistosoma
824	7	0.5	580	10	Q92WP6	Q92wp6 arabidopsis	897	7	0.5	635	2	P71012	P71012 bacillus su
825	7	0.5	580	11	Q98291	Q98291 mus musculus	898	7	0.5	635	5	Q13476	Q13476 homo sapien
826	7	0.5	580	11	Q9EQR3	Q9eqr3 mus musculus	899	7	0.5	637	5	P91656	P91656 drosophila
827	7	0.5	581	5	Q9U4P1	Q9u4p1 drosophila	900	7	0.5	639	3	Q43113	Q43113 neosporea
828	7	0.5	581	10	Q9SX39	Q9sx39 arabidopsis	901	7	0.5	639	4	Q9H8K3	Q9h8k3 homo sapien
829	7	0.5	582	5	Q23909	Q23909 dictyosteli	902	7	0.5	639	10	Q9LN44	Q9ln44 arabidopsis
830	7	0.5	582	5	Q9U4P9	Q9u4p9 drosophila	903	7	0.5	642	2	Q9FC03	Q9fc03 streptomyce
831	7	0.5	582	5	Q9U4P8	Q9u4p8 drosophila	904	7	0.5	642	5	Q93894	Q93894 caenorhabdi
832	7	0.5	582	5	Q9U4P7	Q9u4p7 drosophila	905	7	0.5	646	10	Q9XEC7	Q9xec7 arabidopsis
833	7	0.5	582	5	Q9U4P3	Q9u4p3 drosophila	906	7	0.5	646	11	Q63802	Q63802 rattus norv
834	7	0.5	584	5	Q21088	Q21088 caenorhabdi	907	7	0.5	646	11	Q9EPL7	Q9epL7 mus musculus
835	7	0.5	584	5	Q9U4P6	Q9u4p6 drosophila	908	7	0.5	647	14	Q9DWB9	Q9dwb9 rat cytomeg
836	7	0.5	584	5	Q9U4P5	Q9u4p5 drosophila	909	7	0.5	648	2	Q31644	Q31644 bacillus su
837	7	0.5	584	5	Q9U4P2	Q9u4p2 drosophila	910	7	0.5	649	3	O60167	O60167 schizosacch
838	7	0.5	584	11	Q9EQ61	Q9eq61 mus musculus	911	7	0.5	649	4	Q9F215	Q9f215 homo sapien
839	7	0.5	585	3	Q9HDV2	Q9hdv2 schizosacch	912	7	0.5	650	11	P70663	P70663 mus musculus
840	7	0.5	585	10	Q9LMD2	Q9lmd2 arabidopsis	913	7	0.5	651	5	Q18181	Q18181 caenorhabdi
841	7	0.5	587	2	Q9HUC8	Q9huc8 pseudomonas	914	7	0.5	651	14	Q9WNM8	Q9wnm8 human papil
842	7	0.5	588	5	O16044	O16044 drosophila	915	7	0.5	652	4	Q9Y6N9	Q9y6n9 homo sapien
843	7	0.5	588	5	O9VHJ3	O9vhj3 drosophila	916	7	0.5	654	4	Q9NW82	Q9nw82 homo sapien
844	7	0.5	590	2	Q9KXF5	Q9kyf5 streptomyce	917	7	0.5	656	11	Q61148	Q61148 mus musculus
845	7	0.5	590	4	Q9H2P2	Q9h2p2 homo sapien	918	7	0.5	661	2	Q9JZF3	Q9jzf3 neisseria m
846	7	0.5	592	5	Q9VE48	Q9ve48 drosophila	919	7	0.5	661	5	Q22249	Q22249 caenorhabdi
847	7	0.5	593	11	Q9WU74	Q9wu74 rattus norv	920	7	0.5	662	3	Q9HFB3	Q9hfb3 aspergillus
848	7	0.5	594	11	Q9EQI9	Q9eqi9 mus musculus	921	7	0.5	663	10	Q9XEP8	Q9xep8 sorghum bic
849	7	0.5	594	13	Q9DF69	Q9df69 gallus gall	922	7	0.5	664	2	O05871	O05871 mycobacteri
850	7	0.5	595	4	Q9UEU4	Q9ueu4 homo sapien	923	7	0.5	664	13	Q9IAT6	Q9iat6 brachydanio
851	7	0.5	596	5	O45633	O45633 caenorhabdi	924	7	0.5	667	5	Q9NM60	Q9nm60 caenorhabdi
852	7	0.5	597	4	Q12839	Q12839 homo sapien	925	7	0.5	667	11	O35745	O35745 mus musculus
853	7	0.5	598	4	Q13387	Q13387 homo sapien	926	7	0.5	668	10	Q9SVM5	Q9svm5 arabidopsis
854	7	0.5	599	5	O97009	O97009 leishmania	927	7	0.5	669	4	O14662	O14662 homo sapien
855	7	0.5	600	4	Q9NSB2	Q9nsb2 homo sapien	928	7	0.5	671	4	Q9HAK1	Q9hak1 homo sapien
856	7	0.5	601	3	Q9HE54	Q9he54 neosporea	929	7	0.5	671	4	Q9H4G3	Q9h4g3 homo sapien
857	7	0.5	602	4	Q9HA98	Q9ha98 homo sapien	930	7	0.5	672	5	O44189	O44189 caenorhabdi
858	7	0.5	603	2	Q9EYV2	Q9eyv2 rhizobium l	931	7	0.5	673	5	Q9UON1	Q9uon1 plasmodium
859	7	0.5	603	8	O979097	O979097 alligator m	932	7	0.5	673	11	Q9R0Q1	Q9r0q1 mus musculus
860	7	0.5	603	8	Q33790	Q33790 alligator m	933	7	0.5	674	2	Q9KQ91	Q9kq91 vibrio chol
861	7	0.5	603	10	O22787	O22787 arabidopsis	934	7	0.5	674	4	Q9H053	Q9h053 homo sapien
862	7	0.5	603	10	Q9ST93	Q9st93 oryza sativ	935	7	0.5	675	5	O61310	O61310 trichineila
863	7	0.5	605	3	Q9URJ8	Q9urj8 penicillium	936	7	0.5	675	11	Q9EPT3	Q9ept3 mus musculus
864	7	0.5	605	3	Q9NSK3	Q9nsk3 caenorhabdi	937	7	0.5	676	10	O22762	O22762 arabidopsis
865	7	0.5	607	5	Q9VHA9	Q9vha9 drosophila	938	7	0.5	678	3	O94235	O94235 schizosacch
866	7	0.5	608	5	Q9V6D4	Q9v6d4 drosophila	939	7	0.5	679	2	O916V6	O9i6v6 pseudomonas
867	7	0.5	608	10	Q9SUL2	Q9sul2 arabidopsis	940	7	0.5	680	5	O18496	O18496 caenorhabdi
868	7	0.5	608	11	Q9QVN9	Q9qvn9 cricetus	941	7	0.5	681	5	Q9W8K2	Q9w8k2 drosophila
869	7	0.5	610	3	Q12136	Q12136 saccharomyc	942	7	0.5	682	5	Q9GTW3	Q9gtw3 plasmodium
870	7	0.5	612	5	Q9FXE6	Q9fxe6 arabidopsis	943	7	0.5	683	10	O23506	O23506 arabidopsis
871	7	0.5	614	5	Q9TX06	Q9tx06 dictyosteli	944	7	0.5	684	5	Q9VXS2	Q9vxs2 drosophila
872	7	0.5	614	5	P91956	P91956 lytechinus	945	7	0.5	685	10	Q9SS80	Q9ss80 arabidopsis
873	7	0.5	618	10	Q9M4B7	Q9m4b7 arabidopsis	946	7	0.5	687	11	O89024	O89024 mus musculus
874	7	0.5	618	10	Q9LKL2	Q9lkl2 arabidopsis	947	7	0.5	687	11	Q9WUB6	Q9wub6 mus musculus
875	7	0.5	618	14	Q9QM78	Q9qm78 avian adeno	948	7	0.5	688	11	O88934	O88934 mus musculus
876	7	0.5	619	2	Q9X4G3	Q9x4g3 aquifex pyr	949	7	0.5	691	13	O57658	O57658 gallus gall
877	7	0.5	619	4	O975358	O975358 homo sapien	950	7	0.5	693	14	Q9VQ38	Q9vq38 pseudorabie
878	7	0.5	619	4	Q9HDC6	Q9hdc6 homo sapien	951	7	0.5	694	4	O9Y2J0	O9y2j0 homo sapien
879	7	0.5	619	10	Q9SH81	Q9sh81 arabidopsis	952	7	0.5	694	10	Q9XFS2	Q9xfs2 arabidopsis
880	7	0.5	619	10	Q9LV37	Q9lv37 arabidopsis	953	7	0.5	695	4	O13861	O13861 homo sapien
881	7	0.5	619	14	O73453	O73453 human papil	954	7	0.5	695	11	O64348	O64348 mus musculus
882	7	0.5	621	2	Q9ZFR4	Q9zfr4 burkholderi	955	7	0.5	696	4	O9UJK6	O9ujk6 homo sapien
883	7	0.5	622	13	Q98985	Q98985 rana rugosa	956	7	0.5	696	4	O00461	O00461 homo sapien
884	7	0.5	623	5	Q9W4A2	Q9w4a2 drosophila	957	7	0.5	700	4	O75054	O75054 homo sapien
885	7	0.5	625	2	O67704	O67704 aquifex aeo	958	7	0.5	700	10	Q9SVT1	Q9svt1 arabidopsis
886	7	0.5	626	10	Q9LDL2	Q9ldl2 arabidopsis	959	7	0.5	704	5	Q9VSR3	Q9vsr3 drosophila
887	7	0.5	627	10	Q9MAS7	Q9mas7 arabidopsis	960	7	0.5	704	10	O9FF49	O9ff49 arabidopsis
888	7	0.5	628	2	O84023	O84023 chlamydia t	961	7	0.5	704	11	Q9JLQ4	Q9j1q4 mus musculus
889	7	0.5	629	3	Q9P6U6	Q9p6u6 neosporea	962	7	0.5	708	3	O06411	O06411 saccharomyc
890	7	0.5	629	13	Q93310	Q93310 xenopus lae	963	7	0.5	708	5	O9W1D3	O9w1d3 drosophila
891	7	0.5	630	10	Q9LZE2	Q9lze2 arabidopsis	964	7	0.5	709	5	Q9VU89	Q9vu89 drosophila
892	7	0.5	632	4	Q9NTE0	Q9nte0 homo sapien	965	7	0.5	712	11	O54729	O54729 rattus norv
893	7	0.5	632	4	Q9H5L3	Q9h5l3 homo sapien	966	7	0.5	715	10	Q9LD59	Q9ld59 arabidopsis
894	7	0.5	633	5	O61104	O61104 trypanosoma	967	7	0.5	715	11	Q9Q2X1	Q9q2x1 rattus norv
895	7	0.5	634	4	O15209	O15209 homo sapien	968	7	0.5	719	10	Q40916	Q40916 pseudotsuga



677	7	0.5	477	6	097600	097600 oryctolagus	750	7	0.5	527	10	023389	023389 arabidopsis
678	7	0.5	477	10	09SUL5	09sul5 arabidopsis	751	7	0.5	528	4	095228	095228 homo sapien
679	7	0.5	479	1	09UWR2	09uwr2 pyrococcus	752	7	0.5	528	4	09P0W9	09p0w9 homo sapien
680	7	0.5	480	10	09LSK7	09lsk7 arabidopsis	753	7	0.5	529	10	09FMV0	09fmv0 arabidopsis
681	7	0.5	483	11	063035	063035 rattus norv	754	7	0.5	531	2	09HZ10	09hzi0 pseudomonas
682	7	0.5	484	3	098287	098287 saccharomyc	755	7	0.5	531	5	09TXZ4	09txz4 caenorhabdi
683	7	0.5	484	4	09HC2	09hcg2 homo sapien	756	7	0.5	531	10	09LX13	09lx13 arabidopsis
684	7	0.5	484	10	09SVJ0	09svj0 arabidopsis	757	7	0.5	531	10	09FHT0	09fht0 arabidopsis
685	7	0.5	485	13	09YGS0	09ygs0 gallus gall	758	7	0.5	531	10	09FFQ8	09ffq8 arabidopsis
686	7	0.5	486	3	097376	097376 saccharomyc	759	7	0.5	532	5	09VC95	09vc95 drosophila
687	7	0.5	486	5	09NKW8	09nkw8 patinopecte	760	7	0.5	533	3	001617	001617 caenorhabdi
688	7	0.5	486	10	004591	004591 arabidopsis	761	7	0.5	533	3	006631	006631 saccharomyc
689	7	0.5	486	10	09FRW0	09frw0 arabidopsis	762	7	0.5	536	11	070606	070606 rattus norv
690	7	0.5	487	4	09NQ10	09nql0 homo sapien	763	7	0.5	538	4	09NWQ7	09nwq7 homo sapien
691	7	0.5	487	11	09JLK9	09jlk9 mus musculus	764	7	0.5	538	11	09ET15	09et15 mus musculus
692	7	0.5	490	4	09HBE6	09hbe6 homo sapien	765	7	0.5	539	4	012899	012899 homo sapien
693	7	0.5	491	4	09H4R1	09h4r1 homo sapien	766	7	0.5	539	4	095196	095196 homo sapien
694	7	0.5	491	10	09SKI6	09ski6 arabidopsis	767	7	0.5	539	11	09QX32	09qx32 mus musculus
695	7	0.5	492	6	046384	046384 bos taurus	768	7	0.5	540	4	075864	075864 homo sapien
696	7	0.5	492	11	09Z1K6	09z1k6 mus musculus	769	7	0.5	540	11	062253	062253 mus musculus
697	7	0.5	493	2	09JMM8	09jmm8 wolbachia s	770	7	0.5	541	3	09P6T4	09p6t4 neuospora
698	7	0.5	493	4	095376	095376 homo sapien	771	7	0.5	544	4	09P0J8	09p0j8 homo sapien
699	7	0.5	493	4	09UEW9	09uem9 homo sapien	772	7	0.5	544	11	062831	062831 rattus norv
700	7	0.5	493	10	093015	093015 arabidopsis	773	7	0.5	545	3	095842	095842 saccharomyc
701	7	0.5	493	10	09SKS9	09sk9 arabidopsis	774	7	0.5	545	6	077666	077666 sus scrofa
702	7	0.5	494	3	09HEL2	09hel2 neuospora	775	7	0.5	545	10	09M9V8	09m9v8 arabidopsis
703	7	0.5	495	10	023330	023330 arabidopsis	776	7	0.5	545	11	088836	088836 mus musculus
704	7	0.5	496	4	09Y2K7	09y2k7 homo sapien	777	7	0.5	547	10	09SD74	09sd74 arabidopsis
705	7	0.5	497	11	09J110	09j110 mus musculus	778	7	0.5	548	11	09ES65	09es65 mus musculus
706	7	0.5	498	10	093483	093483 calocedrus	779	7	0.5	549	5	09N5M7	09n5m7 caenorhabdi
707	7	0.5	498	10	09LZC8	09lzc8 arabidopsis	780	7	0.5	549	13	09DGM7	09dgm7 ictalurus p
708	7	0.5	499	5	09Y120	09y120 drosophila	781	7	0.5	550	5	09VD69	09vd69 drosophila
709	7	0.5	499	10	09FWY3	09fwy3 lycopersico	782	7	0.5	550	10	082347	082347 arabidopsis
710	7	0.5	500	4	09UN16	09un16 homo sapien	783	7	0.5	551	4	013512	013512 homo sapien
711	7	0.5	500	5	022469	022469 caenorhabdi	784	7	0.5	551	4	015740	015740 homo sapien
712	7	0.5	501	2	09ZJ03	09zj03 streptococc	785	7	0.5	551	13	013003	013003 gallus gall
713	7	0.5	502	10	09LXH2	09lyh2 arabidopsis	786	7	0.5	552	2	067593	067593 aquifex aeo
714	7	0.5	502	10	065237	065237 arabidopsis	787	7	0.5	552	2	09K5M6	09k5m6 anabaena sp
715	7	0.5	502	11	063892	063892 rattus norv	788	7	0.5	552	4	09UM04	09um04 homo sapien
716	7	0.5	507	11	09RQ00	09rq00 mus musculus	789	7	0.5	552	5	046178	046178 strongyloce
717	7	0.5	503	5	09VL48	09vl48 drosophila	790	7	0.5	552	5	045117	045117 chironomus
718	7	0.5	504	5	09VFW0	09vfm0 drosophila	791	7	0.5	552	10	09LIS7	09lis7 oryza sativ
719	7	0.5	504	10	09SUE9	09sue9 arabidopsis	792	7	0.5	553	10	049931	049931 pisum sativ
720	7	0.5	505	11	060569	060569 mus cookii.	793	7	0.5	554	11	09JHT9	09jht9 mus musculus
721	7	0.5	507	11	054716	054716 mus musculus	794	7	0.5	558	5	026631	026631 strongyloce
722	7	0.5	507	11	035728	035728 mus musculus	795	7	0.5	558	10	09M8S2	09m8s2 arabidopsis
723	7	0.5	508	4	015208	015208 homo sapien	796	7	0.5	559	10	09M016	09m016 arabidopsis
724	7	0.5	508	10	093484	093484 calocedrus	797	7	0.5	560	4	09NY90	09ny90 homo sapien
725	7	0.5	509	10	040870	040870 picea glauc	798	7	0.5	560	6	046411	046411 bos taurus
726	7	0.5	510	4	075926	075926 homo sapien	799	7	0.5	560	10	064883	064883 arabidopsis
727	7	0.5	514	4	015223	015223 homo sapien	800	7	0.5	562	4	012820	012820 homo sapien
728	7	0.5	515	5	09V8M3	09v8m3 drosophila	801	7	0.5	564	5	09VRX8	09vrx8 drosophila
729	7	0.5	515	6	09GL76	09gl76 sus scrofa	802	7	0.5	564	5	09V719	09v719 drosophila
730	7	0.5	515	11	09JEF6	09jef6 mus musculus	803	7	0.5	565	4	09UB01	09ubq1 homo sapien
731	7	0.5	515	11	09ERL5	09erl5 mus musculus	804	7	0.5	565	4	09NZS5	09nzs5 homo sapien
732	7	0.5	516	10	09LZS1	09lzs1 arabidopsis	805	7	0.5	565	5	090625	090625 cryptospori
733	7	0.5	516	11	09J117	09j117 mus musculus	806	7	0.5	566	6	09TU19	09t19 canis famli
734	7	0.5	517	4	075465	075465 homo sapien	807	7	0.5	566	5	09W243	09w243 drosophila
735	7	0.5	517	5	017941	017941 caenorhabdi	808	7	0.5	566	10	024294	024294 pisum sativ
736	7	0.5	517	5	09V921	09v921 drosophila	809	7	0.5	567	4	09HD28	09hd28 homo sapien
737	7	0.5	519	4	09Y508	09y508 homo sapien	810	7	0.5	567	4	09H606	09h6q6 homo sapien
738	7	0.5	519	4	09H4P2	09h4p2 homo sapien	811	7	0.5	567	4	09H606	09h6q6 mus musculus
739	7	0.5	521	4	09UM17	09um17 homo sapien	812	7	0.5	569	10	043468	043468 glycine max
740	7	0.5	522	10	09L1W0	09liw0 oryza sativ	813	7	0.5	571	11	09ERQ6	09erq6 rattus norv
741	7	0.5	523	3	008831	008831 saccharomyc	814	7	0.5	572	5	020715	020715 caenorhabdi
742	7	0.5	523	10	09LJZ2	09ljz2 arabidopsis	815	7	0.5	573	4	09P2A5	09p2a5 homo sapien
743	7	0.5	523	13	093482	093482 gallus gall	816	7	0.5	574	4	09NXT6	09nxt6 homo sapien
744	7	0.5	524	10	09LME6	09lme6 arabidopsis	817	7	0.5	574	4	09H4L4	09h4l4 homo sapien
745	7	0.5	524	10	09FJY5	09fjy5 arabidopsis	818	7	0.5	574	11	09WU75	09wu75 rattus norv
746	7	0.5	525	11	09WU76	09wu76 rattus norv	819	7	0.5	576	10	09LUY5	09luys arabidopsis
747	7	0.5	526	2	09RQ57	09rq57 mycoplasma	820	7	0.5	578	5	09U4P4	09u4p4 drosophila
748	7	0.5	526	2	091365	091365 pseudomonas	821	7	0.5	580	4	09UF71	09uf71 homo sapien
749	7	0.5	526	4	09UP59	09up59 homo sapien	822	7	0.5	580	5	09TW69	09tw69 drosophila



531	7	0.5	360	14	039839	039839 influenza a	604	7	0.5	415	4	095906	095906 homo sapien
532	7	0.5	363	2	025617	025617 helicobacte	605	7	0.5	415	4	09NRD5	09nrD5 homo sapien
533	7	0.5	367	4	09P018	09P018 homo sapien	606	7	0.5	416	10	09SVG6	09svG6 arabisopsis
534	7	0.5	368	10	09LIC1	09lic1 arabisopsis	607	7	0.5	417	2	09WZ24	09wZ24 thermotoga
535	7	0.5	368	10	09S7Y1	09S7Y1 arabisopsis	608	7	0.5	420	4	014979	014979 homo sapien
536	7	0.5	369	2	09JW8	09Jw8 neisseria m	609	7	0.5	420	4	09NSR2	09nsR2 homo sapien
537	7	0.5	370	11	035987	035987 rattus norv	610	7	0.5	420	10	09SJF6	09sJf6 arabisopsis
538	7	0.5	371	5	09VTS9	09vts9 drosophila	611	7	0.5	421	5	09U6S0	09u6S0 strongyloce
539	7	0.5	372	2	074455	074455 synecocyst	612	7	0.5	422	5	09W4E5	09w4E5 drosophila
540	7	0.5	372	10	080888	080888 arabisopsis	613	7	0.5	422	14	092503	092503 bombyx mori
541	7	0.5	373	4	09UQ42	09uq42 homo sapien	614	7	0.5	423	10	09LQO8	09lqQ8 arabisopsis
542	7	0.5	373	4	09Y2Y7	09y2Y7 homo sapien	615	7	0.5	424	5	017469	017469 caenorhabditi
543	7	0.5	373	5	026960	026960 trypanosoma	616	7	0.5	424	10	09ZNR7	09znR7 arabisopsis
544	7	0.5	374	10	09LR97	09lr97 arabisopsis	617	7	0.5	426	5	09NF73	09nf73 drosophila
545	7	0.5	374	5	09VVB5	09vVB5 drosophila	618	7	0.5	427	3	074682	074682 aspergillus
546	7	0.5	374	14	09WF15	09wF15 human adeno	619	7	0.5	427	11	064321	064321 mus musculu
547	7	0.5	376	2	09ZC09	09zc09 streptomyces	620	7	0.5	429	3	09P6R2	09p6R2 schizosacch
548	7	0.5	377	11	09Z223	09z223 cricetus	621	7	0.5	429	10	09SAD0	09sAd0 arabisopsis
549	7	0.5	378	5	09VRR9	09vrr9 drosophila	622	7	0.5	431	1	027685	027685 methanobact
550	7	0.5	379	5	025705	025705 plasmodium	623	7	0.5	431	10	09FLP9	09flP9 arabisopsis
551	7	0.5	379	5	025706	025706 plasmodium	624	7	0.5	433	2	09RTF6	09rtf6 deinococcus
552	7	0.5	379	5	09U6C4	09u6C4 plasmodium	625	7	0.5	434	10	065310	065310 arabisopsis
553	7	0.5	379	10	09W8S5	09w8S5 arabisopsis	626	7	0.5	434	10	09SBJ5	09sBj5 arabisopsis
554	7	0.5	380	5	026019	026019 plasmodium	627	7	0.5	435	10	022281	022281 arabisopsis
555	7	0.5	380	5	09XZ72	09xz72 libellula p	628	7	0.5	435	10	09FKN6	09fkn6 arabisopsis
556	7	0.5	381	3	09P625	09p625 neurospora	629	7	0.5	436	3	036020	036020 schizosacch
557	7	0.5	381	4	09UQC6	09uqc6 homo sapien	630	7	0.5	436	4	043440	043440 homo sapien
558	7	0.5	382	5	09GQ81	09gq81 drosophila	631	7	0.5	438	11	055036	055036 cricetus
559	7	0.5	382	8	099971	099971 porphyra pu	632	7	0.5	438	11	09ER75	09er75 mus musculu
560	7	0.5	382	13	09P5W8	09p5W8 xenopus lae	633	7	0.5	439	2	09L516	09l516 salmonella
561	7	0.5	383	11	09WVM2	09wVm2 mus musculu	634	7	0.5	439	5	09GTH6	09gth6 caenorhabditi
562	7	0.5	384	5	09XZ71	09xz71 periplaneta	635	7	0.5	440	4	09NV78	09nv78 homo sapien
563	7	0.5	384	5	09V394	09v394 drosophila	636	7	0.5	441	4	09NW01	09nw01 homo sapien
564	7	0.5	386	3	09P7B0	09p7B0 schizosacch	637	7	0.5	441	10	09STC9	09stc9 zinnia eleg
565	7	0.5	386	5	044534	044534 aedes aegypt	638	7	0.5	443	10	09ZR26	09zr26 perilla fru
566	7	0.5	386	13	091839	091839 xenopus lae	639	7	0.5	443	13	0918W3	0918W3 gallus gall
567	7	0.5	386	14	09IMN4	09imN4 hydrangea m	640	7	0.5	444	4	09UPW2	09upW2 homo sapien
568	7	0.5	387	2	09I296	09i296 pseudomonas	641	7	0.5	445	5	09VN84	09vn84 drosophila
569	7	0.5	388	5	09VVB3	09vVB3 drosophila	642	7	0.5	445	10	09ZWC5	09zwc5 arabisopsis
570	7	0.5	388	10	048678	048678 arabisopsis	643	7	0.5	445	10	09M1Y5	09m1Y5 arabisopsis
571	7	0.5	388	13	09DF24	09df24 xenopus lae	644	7	0.5	445	11	060877	060877 mus musculu
572	7	0.5	389	6	09GL07	09gl07 phoca vitul	645	7	0.5	445	11	035371	035371 mus musculu
573	7	0.5	391	5	09VTT2	09vtt2 drosophila	646	7	0.5	447	4	09H931	09h931 homo sapien
574	7	0.5	391	6	09GL06	09gl06 sus scrofa	647	7	0.5	448	4	09Y2V1	09y2V1 homo sapien
575	7	0.5	391	10	09L127	09l127 oryza sativ	648	7	0.5	448	6	09XS63	09xs63 equus cabal
576	7	0.5	392	4	013149	013149 homo sapien	649	7	0.5	448	10	09STM4	09stm4 arabisopsis
577	7	0.5	393	3	013701	013701 schizosacch	650	7	0.5	449	4	09UEI6	09ueI6 homo sapien
578	7	0.5	393	6	09GL16	09gl16 hippopotamu	651	7	0.5	450	10	09SUC3	09suc3 arabisopsis
579	7	0.5	393	6	09GL15	09gl15 macrotus ca	652	7	0.5	452	5	019175	019175 caenorhabditi
580	7	0.5	393	6	09GL11	09gl11 nycticebus	653	7	0.5	452	10	09SGX8	09sgX8 arabisopsis
581	7	0.5	393	11	070278	070278 mus musculu	654	7	0.5	454	10	004500	004500 arabisopsis
582	7	0.5	394	1	09UYC0	09uyC0 pyrococcus	655	7	0.5	454	11	09QY91	09qy91 mus musculu
583	7	0.5	395	6	09GL12	09gl12 manis sp. a	656	7	0.5	455	3	012080	012080 saccharomyc
584	7	0.5	395	13	042481	042481 makaira nig	657	7	0.5	455	11	062511	062511 mus musculu
585	7	0.5	396	2	09F793	09f793 campylobact	658	7	0.5	456	10	09SL95	09sl95 arabisopsis
586	7	0.5	396	11	09Z1K1	09z1k1 mus musculu	659	7	0.5	457	4	09NSD2	09nsd2 homo sapien
587	7	0.5	397	5	09VVB4	09vVB4 drosophila	660	7	0.5	457	4	09HAK6	09hak6 homo sapien
588	7	0.5	398	2	052121	052121 escherichia	661	7	0.5	460	4	09ULW6	09ulW6 homo sapien
589	7	0.5	399	4	09NYB0	09nyB0 homo sapien	662	7	0.5	461	2	09I6E1	09i6E1 pseudomonas
590	7	0.5	399	10	09LW33	09lW33 arabisopsis	663	7	0.5	461	10	09ZRG9	09zrg9 oryza sativ
591	7	0.5	400	5	027934	027934 drosophila	664	7	0.5	461	10	09FGZ7	09fgZ7 arabisopsis
592	7	0.5	400	10	09ZRM3	09zrm3 antirrhinum	665	7	0.5	463	2	09RNM3	09rnm3 zymomonas m
593	7	0.5	405	4	013215	013215 homo sapien	666	7	0.5	464	4	043159	043159 homo sapien
594	7	0.5	405	5	026268	026268 aplysia cal	667	7	0.5	466	5	093712	093712 caenorhabditi
595	7	0.5	405	11	09WV69	09wv69 mus musculu	668	7	0.5	469	1	072021	072021 methanosarc
596	7	0.5	407	5	09TX81	09tx81 plasmodium	669	7	0.5	471	2	09HWN7	09hwn7 pseudomonas
597	7	0.5	407	10	049747	049747 arabisopsis	670	7	0.5	471	4	09H0V2	09h0V2 homo sapien
598	7	0.5	409	10	09SGC6	09sgc6 arabisopsis	671	7	0.5	471	10	09LXJ8	09lXj8 arabisopsis
599	7	0.5	409	10	09LXJ8	09lXJ8 arabisopsis	672	7	0.5	472	1	09VZ24	09vZ24 methanosarc
600	7	0.5	412	4	013558	013558 homo sapien	673	7	0.5	473	13	073840	073840 gallus gall
601	7	0.5	412	4	09ULX3	09ulX3 homo sapien	674	7	0.5	474	4	09UK66	09uk66 homo sapien
602	7	0.5	412	10	09FZC9	09fzc9 arabisopsis	675	7	0.5	475	1	058382	058382 pyrococcus
603	7	0.5	414	3	09P5H7	09p5H7 neurospora	676	7	0.5	476	4	09UJ08	09uJ08 homo sapien



385	7	0.5	270	5	096536	096536 strongyloce	458	7	0.5	316	10	09FX02	09fx02 arabidopsis
386	7	0.5	270	5	09TZ06	Q9tzq6 strongyloce	459	7	0.5	316	11	Q62254	Q62254 mus musculus
387	7	0.5	271	2	09KUK5	Q9kuk5 vibrio chol	460	7	0.5	316	11	P70113	P70113 mesocricetu
388	7	0.5	271	10	09SV06	Q9sv06 arabidopsis	461	7	0.5	320	4	Q9H4R0	Q9h4r0 homo sapien
389	7	0.5	273	10	09FHB2	Q9fhb2 arabidopsis	462	7	0.5	320	6	P79119	P79119 bos taurus
390	7	0.5	275	13	013090	013090 pleurodeles	463	7	0.5	322	4	Q99645	Q99645 homo sapien
391	7	0.5	276	2	09F733	Q9f733 chlorobium	464	7	0.5	323	2	P95185	P95185 mycobacteri
392	7	0.5	276	13	P79780	P79780 gallus gall	465	7	0.5	323	13	073907	073907 gallus gall
393	7	0.5	279	10	09FHY1	Q9fhy1 arabidopsis	466	7	0.5	324	2	Q92834	Q92834 chlamydia p
394	7	0.5	280	4	09NVW5	Q9nvw5 homo sapien	467	7	0.5	324	3	Q9HEF2	Q9hef2 neurospora
395	7	0.5	283	13	09I939	Q9i939 misgurnus a	468	7	0.5	326	4	Q9HAJ4	Q9haj4 homo sapien
396	7	0.5	284	4	Q9H215	Q9h215 homo sapien	469	7	0.5	326	5	Q9VPX6	Q9vpix6 drosophila
397	7	0.5	285	11	Q9Z2D8	Q9z2d8 mus musculus	470	7	0.5	328	10	Q9ZVX7	Q9zvx7 arabidopsis
398	7	0.5	286	2	Q50196	Q50196 mycobacteri	471	7	0.5	328	13	Q9PRJ2	Q9prj2 brachydanio
399	7	0.5	286	14	Q9DH21	Q9dh21 tt virus. v	472	7	0.5	329	2	084436	084436 chlamydia t
400	7	0.5	288	5	017375	017375 caenorhabdi	473	7	0.5	329	4	Q9NWV8	Q9nwv8 homo sapien
401	7	0.5	289	11	Q9JK23	Q9jk23 mus musculus	474	7	0.5	329	5	Q9NFV9	Q9nf9v plasmodium
402	7	0.5	290	14	Q9Q8V0	Q9q8v0 shope fibro	475	7	0.5	329	10	Q9FH65	Q9fh65 arabidopsis
403	7	0.5	291	2	Q9S0F2	Q9s0f2 borrelia bu	476	7	0.5	330	6	Q9GLY8	Q9gly8 macaca fasc
404	7	0.5	291	2	Q9S0C3	Q9s0c3 borrelia bu	477	7	0.5	331	2	Q9SLJ8	Q9slj8 streptococc
405	7	0.5	291	2	Q9S071	Q9s071 borrelia bu	478	7	0.5	331	2	Q9FDD3	Q9fdd3 brucella ab
406	7	0.5	291	2	Q9S046	Q9s046 borrelia bu	479	7	0.5	331	3	Q9Y7T2	Q9y7t2 schizosacch
407	7	0.5	291	2	Q9RZ25	Q9rzz5 borrelia bu	480	7	0.5	331	4	Q9HD90	Q9hd90 homo sapien
408	7	0.5	291	2	Q9R331	Q9r331 borrelia bu	481	7	0.5	331	10	Q9FX56	Q9fx56 arabidopsis
409	7	0.5	291	4	Q9S983	Q9s983 homo sapien	482	7	0.5	332	2	Q9PJW3	Q9pjw3 chlamydia m
410	7	0.5	291	4	Q9NUL5	Q9nul5 homo sapien	483	7	0.5	332	10	Q9LQW2	Q9lqw2 arabidopsis
411	7	0.5	292	2	044793	Q44793 borrelia bu	484	7	0.5	335	10	Q9S9Q6	Q9s9q6 arabidopsis
412	7	0.5	292	2	Q9X217	Q9x217 thermotoga	485	7	0.5	337	5	Q9Y156	Q9y156 drosophila
413	7	0.5	292	2	Q9S088	Q9s088 borrelia bu	486	7	0.5	338	14	Q9DM07	Q9dm07 plutella xy
414	7	0.5	292	2	Q9L9N7	Q9l9n7 borrelia bu	487	7	0.5	339	10	Q9LQW4	Q9lqw4 arabidopsis
415	7	0.5	292	10	Q48730	Q48730 arabidopsis	488	7	0.5	339	11	Q61149	Q61149 mus musculus
416	7	0.5	292	10	Q9Z090	Q9zu90 arabidopsis	489	7	0.5	340	2	Q9KN50	Q9kns0 vibrio chol
417	7	0.5	293	4	Q9UDC3	Q9udc3 homo sapien	490	7	0.5	340	4	Q9NXZ4	Q9nxz4 homo sapien
418	7	0.5	293	10	Q9L179	Q9l179 arabidopsis	491	7	0.5	340	5	Q9U7C9	Q9u7c9 dictyosteli
419	7	0.5	293	10	Q9FJ35	Q9fj35 arabidopsis	492	7	0.5	340	10	Q9LFS1	Q9lfs1 arabidopsis
420	7	0.5	293	13	Q9I8M8	Q9i8md petromyzon	493	7	0.5	341	3	074509	074509 schizosacch
421	7	0.5	293	14	Q91824	Q91824 coho salmon	494	7	0.5	342	5	Q9NGS6	Q9ngs6 acropora mi
422	7	0.5	294	14	Q9QF47	Q9qf47 avian pneum	495	7	0.5	343	10	064878	064878 arabidopsis
423	7	0.5	295	10	Q9ZW81	Q9zwm1 arabidopsis	496	7	0.5	344	11	Q9ER45	Q9er45 mus musculus
424	7	0.5	295	10	Q9LSZ8	Q9lsz8 arabidopsis	497	7	0.5	346	5	Q9U0G0	Q9u0g0 plasmodium
425	7	0.5	296	2	Q9Z8S5	Q9z8s5 chlamydia p	498	7	0.5	347	4	Q9UKP3	Q9ukp3 homo sapien
426	7	0.5	297	5	Q9VEM3	Q9vem3 drosophila	499	7	0.5	347	4	Q9HAF9	Q9haf9 homo sapien
427	7	0.5	298	1	Q9S8205	Q9s8205 pyrococcus	500	7	0.5	347	5	061872	061872 caenorhabdi
428	7	0.5	298	3	Q9S9842	Q9s9842 penicillium	501	7	0.5	347	13	Q9DE43	Q9ded3 brachydanio
429	7	0.5	299	1	Q9UYE4	Q9uye4 pyrococcus	502	7	0.5	348	10	Q9LU06	Q9lu06 arabidopsis
430	7	0.5	299	5	Q9VWL3	Q9vwl3 drosophila	503	7	0.5	348	10	Q9FLX8	Q9flx8 arabidopsis
431	7	0.5	299	10	082801	082801 arabidopsis	504	7	0.5	349	10	Q9LU07	Q9lu07 arabidopsis
432	7	0.5	299	10	048722	048722 arabidopsis	505	7	0.5	349	13	Q9YHY7	Q9yhy7 gallus gall
433	7	0.5	299	11	Q9QXU1	Q9qyu1 rattus ratt	506	7	0.5	350	4	Q9S455	Q9s455 homo sapien
434	7	0.5	299	11	Q9OQU1	Q9quq1 mus musculu	507	7	0.5	350	5	076720	076720 caenorhabdi
435	7	0.5	300	6	046383	046383 bison bison	508	7	0.5	350	5	Q9NE21	Q9ne21 leishmania
436	7	0.5	301	2	Q9RGF0	Q9rgf0 campylobact	509	7	0.5	350	10	Q9SMQ4	Q9smq4 arabidopsis
437	7	0.5	301	2	045333	045333 caenorhabdi	510	7	0.5	350	13	042202	042202 brachydanio
438	7	0.5	301	11	Q9Z130	Q9z130 mus musculus	511	7	0.5	351	4	Q9ULR0	Q9ulr0 homo sapien
439	7	0.5	302	10	Q9FX68	Q9fx68 arabidopsis	512	7	0.5	351	10	Q9ZVU8	Q9zvu8 arabidopsis
440	7	0.5	302	10	Q9FN16	Q9fn16 arabidopsis	513	7	0.5	351	10	Q9M4F3	Q9m4f3 brassica na
441	7	0.5	303	9	Q9ZX05	Q9zxx05 mycobacteri	514	7	0.5	351	10	Q9FX55	Q9fx55 arabidopsis
442	7	0.5	303	10	Q9M0H2	Q9m0h2 arabidopsis	515	7	0.5	352	11	Q9R192	Q9r192 rattus norv
443	7	0.5	303	10	Q9LYP4	Q9lyp4 arabidopsis	516	7	0.5	353	4	Q9UN41	Q9un41 homo sapien
444	7	0.5	304	10	Q9FRA5	Q9fra5 oryza sativ	517	7	0.5	353	10	Q9SAL5	Q9sal5 arabidopsis
445	7	0.5	306	2	067883	067883 aquifex aeo	518	7	0.5	353	10	Q9LVN5	Q9lvn5 arabidopsis
446	7	0.5	306	5	017377	017377 caenorhabdi	519	7	0.5	354	5	Q25995	Q25995 plasmodium
447	7	0.5	307	2	052061	052061 bacillus me	520	7	0.5	354	10	Q9FK12	Q9fk12 arabidopsis
448	7	0.5	308	4	Q9HAY2	Q9hay2 homo sapien	521	7	0.5	354	11	062509	062509 mus musculus
449	7	0.5	310	10	Q9LTP9	Q9ltb9 arabidopsis	522	7	0.5	355	1	Q9UX70	Q9ux70 sulfolobus
450	7	0.5	314	1	Q9HJM0	Q9hjm0 thermoplasm	523	7	0.5	355	2	Q9L1A1	Q9l1al streptomyce
451	7	0.5	314	5	Q20146	Q20146 caenorhabdi	524	7	0.5	356	4	Q9UEC8	Q9uec8 homo sapien
452	7	0.5	314	11	Q9ESH2	Q9esh2 rattus norv	525	7	0.5	356	11	Q9R193	Q9r193 rattus norv
453	7	0.5	315	4	Q9NUD4	Q9nud4 homo sapien	526	7	0.5	357	2	083906	083906 treponema p
454	7	0.5	315	5	044605	044605 caenorhabdi	527	7	0.5	357	4	Q9UN42	Q9un42 homo sapien
455	7	0.5	316	2	P75300	P75300 mycoplasma	528	7	0.5	359	5	Q9YOC6	Q9yoc6 trichinella
456	7	0.5	316	10	Q42866	Q42866 lupinus alb	529	7	0.5	359	5	Q9XV41	Q9xv41 caenorhabdi
457	7	0.5	316	10	Q9MAA2	Q9maa2 arabidopsis	530	7	0.5	360	10	Q9FFP9	Q9ffp9 arabidopsis



239	7	0.5	199	14	Q9W9X3	Q9W9X3 porcine ade	312	7	0.5	235	5	096535	096535 strongyloce
240	7	0.5	200	4	O15347	Q15347 homo sapien	313	7	0.5	235	10	Q9L1Z2	Q9L1Z2 oryza sativ
241	7	0.5	200	10	O23294	O23294 arabisopsis	314	7	0.5	236	2	Q9R5Z7	Q9R5Z7 bacteroides
242	7	0.5	200	11	O54879	O54879 mus musculu	315	7	0.5	236	5	O24872	O24872 echinometra
243	7	0.5	201	4	O95556	O95556 homo sapien	316	7	0.5	236	5	O24874	O24874 echinometra
244	7	0.5	202	2	Q9F0Z7	Q9F0Z7 bradyrhizob	317	7	0.5	236	5	O24875	O24875 echinometra
245	7	0.5	203	3	Q9P7J2	Q9P7J2 schizosacch	318	7	0.5	236	5	096538	096538 hemicentrot
246	7	0.5	204	10	Q92954	P92954 arabisopsis	319	7	0.5	236	5	Q9T2Q8	Q9T2Q8 strongyloce
247	7	0.5	204	10	O43386	O43386 arabisopsis	320	7	0.5	236	5	Q9T2Q7	Q9T2Q7 strongyloce
248	7	0.5	205	2	O31812	O31812 bacillus su	321	7	0.5	236	5	Q9T2Q5	Q9T2Q5 hemicentrot
249	7	0.5	206	2	P96510	P96510 borrelia af	322	7	0.5	236	10	Q9L1W6	Q9L1W6 oryza sativ
250	7	0.5	206	2	Q9L6N7	Q9L6N7 salmonella	323	7	0.5	238	4	Q9H032	Q9H032 homo sapien
251	7	0.5	208	2	O57530	Q57530 agrobacteri	324	7	0.5	238	10	Q9FNL9	Q9FNL9 arabisopsis
252	7	0.5	208	10	Q9T089	Q9T089 arabisopsis	325	7	0.5	238	14	Q9WFX9	Q9WFX9 human herpe
253	7	0.5	210	4	O95054	Q95054 homo sapien	326	7	0.5	238	14	Q9WFX8	Q9WFX8 human herpe
254	7	0.5	211	11	Q9EQD5	O9eqd5 mus musculu	327	7	0.5	238	14	Q9W891	Q9W891 human herpe
255	7	0.5	211	2	O9Z357	Q9Z357 borrelia va	328	7	0.5	239	6	Q9XSK7	Q9XSK7 bos taurus
256	7	0.5	211	5	Q9NAA6	Q9naa6 caenorhabdi	329	7	0.5	239	14	Q9W8W3	Q9W8W3 human herpe
257	7	0.5	213	11	Q9JHX0	Q9jhx0 rattus norv	330	7	0.5	241	5	O24907	O24907 echinometra
258	7	0.5	213	4	Q9UJG1	Q9ujg1 homo sapien	331	7	0.5	241	5	O24906	O24906 echinometra
259	7	0.5	213	10	O43600	O43600 oryza sativ	332	7	0.5	241	5	O24905	O24905 echinometra
260	7	0.5	213	10	Q9LPM6	Q9lpw6 arabisopsis	333	7	0.5	241	10	Q9ZQO5	Q9ZQO5 arabisopsis
261	7	0.5	214	1	O23215	O23215 archaeoglob	334	7	0.5	242	5	O24909	Q24909 echinometra
262	7	0.5	214	3	O42932	O42932 schizosacch	335	7	0.5	242	10	Q9XE71	Q9XE71 sorghum bic
263	7	0.5	214	10	O49780	O49780 arabisopsis	336	7	0.5	242	10	Q9LDX4	Q9Ldx4 oryza sativ
264	7	0.5	214	13	Q9PEUK9	Q9pek9 gallus gall	337	7	0.5	242	4	Q9NKK9	Q9nkk9 homo sapien
265	7	0.5	215	2	O92EJ9	Q9zef9 borrelia af	338	7	0.5	243	5	Q9H1T9	Q9hit9 homo sapien
266	7	0.5	215	2	P70810	P70810 borrelia af	339	7	0.5	243	5	O24908	O24908 echinometra
267	7	0.5	215	10	O65415	O65415 arabisopsis	340	7	0.5	245	10	Q9FVE6	Q9fve6 arabisopsis
268	7	0.5	215	11	O88611	O88611 spalax leuc	341	7	0.5	245	5	Q9W3H4	Q9w3h4 drosophila
269	7	0.5	215	13	Q9YH06	Q9yh06 gallus gall	342	7	0.5	247	10	O81766	O81766 arabisopsis
270	7	0.5	215	14	Q9QTF1	Q9qtf1 unidentified	343	7	0.5	247	10	O81766	O81766 arabisopsis
271	7	0.5	216	6	Q9N213	Q9n213 macaca sylv	344	7	0.5	248	5	O25676	O25676 plasmodium
272	7	0.5	216	6	Q9N209	Q9n209 macaca sylv	345	7	0.5	248	10	O80892	O80892 arabisopsis
273	7	0.5	216	6	Q9N207	Q9n207 callithrix	346	7	0.5	248	10	O49693	O49693 arabisopsis
274	7	0.5	216	10	Q9SLG7	Q9slg7 arabisopsis	347	7	0.5	249	5	Q9UIN3	Q9uin3 cryptospori
275	7	0.5	216	11	O08904	O08904 mus musculu	348	7	0.5	249	9	O80305	O80305 bacterioph
276	7	0.5	217	3	Q9HEX7	Q9hex7 pneumocysti	349	7	0.5	249	10	O22952	O22952 arabisopsis
277	7	0.5	217	4	Q9H7Y2	Q9h7y2 homo sapien	350	7	0.5	250	2	Q9ZDM4	Q9zdm4 rickettsia
278	7	0.5	217	5	O97461	O97461 strongyloce	351	7	0.5	250	5	O62387	Q62387 caenorhabdi
279	7	0.5	217	10	Q9LYL1	Q9lyl1 arabisopsis	352	7	0.5	250	5	O24917	Q24917 echinometra
280	7	0.5	217	13	Q9IB51	Q9ib51 oncorhynch	353	7	0.5	250	11	O88795	O88795 mus musculu
281	7	0.5	219	5	O96537	O96537 allocentrot	354	7	0.5	251	5	O24916	O24916 echinometra
282	7	0.5	219	5	O9TYD4	Q9tyd4 allocentrot	355	7	0.5	251	13	Q9VI93	Q9vi93 gallus gall
283	7	0.5	221	10	Q9ZVR9	Q9zvr9 arabisopsis	356	7	0.5	254	10	O48767	O48767 arabisopsis
284	7	0.5	223	4	Q9NQF9	Q9ncf9 homo sapien	357	7	0.5	255	4	Q9NWB0	Q9nwb0 homo sapien
285	7	0.5	223	10	Q9M2Y1	Q9m2y1 arabisopsis	358	7	0.5	255	2	P95273	P95273 mycobacteri
286	7	0.5	224	2	O55322	O55322 synechococc	359	7	0.5	256	10	Q9FSJ4	Q9fsj4 oryza sativ
287	7	0.5	224	4	O94797	O94797 homo sapien	360	7	0.5	256	10	Q9P1V9	Q9p1v9 homo sapien
288	7	0.5	228	1	O29479	O29479 archaeoglob	361	7	0.5	257	10	O22238	O22238 arabisopsis
289	7	0.5	228	4	O9P0A8	Q9p0a8 homo sapien	362	7	0.5	259	2	Q9Z9V7	Q9z9v7 bacillus ha
290	7	0.5	228	5	O96534	O96534 strongyloce	363	7	0.5	259	2	Q9KGA5	Q9kga5 bacillus ha
291	7	0.5	229	10	O24480	O24480 allium porr	364	7	0.5	259	13	Q91964	Q91964 gallus gall
292	7	0.5	230	3	Q9URZ1	Q9urz1 schizosacch	365	7	0.5	260	5	Q9XUN9	Q9xun9 caenorhabdi
293	7	0.5	230	5	O24870	O24870 echinometra	366	7	0.5	260	11	P97822	P97822 mus musculu
294	7	0.5	232	1	O28269	O28269 archaeoglob	367	7	0.5	261	5	Q9NF82	Q9nf82 leishmania
295	7	0.5	232	5	O24871	O24871 echinometra	368	7	0.5	261	10	Q9SK91	Q9sk91 arabisopsis
296	7	0.5	232	10	Q9ST20	Q9st20 brassica ca	369	7	0.5	262	2	Q91AHL	Q91ahl pseudomonas
297	7	0.5	233	2	Q9HTY5	Q9hcy5 pseudomonas	370	7	0.5	262	3	O74515	O74515 schizosacch
298	7	0.5	233	5	O44377	O44377 arabia pun	371	7	0.5	262	3	O74515	O74515 schizosacch
299	7	0.5	233	5	O46361	O46361 arabia pun	372	7	0.5	262	14	Q9ULM8	Q9ulm8 homo sapien
300	7	0.5	234	6	Q9T530	Q9ts30 oryctolag	373	7	0.5	262	14	O22547	O22547 epizootic h
301	7	0.5	235	2	O9PNC3	Q9pnc3 campylobact	374	7	0.5	264	10	O04470	O04470 arabisopsis
302	7	0.5	235	5	O44373	O44373 arabia lix	375	7	0.5	265	2	Q9KCA9	Q9kca9 bacillus ha
303	7	0.5	235	5	O44374	O44374 arabia lix	376	7	0.5	265	10	Q9M2Q7	Q9m2q7 arabisopsis
304	7	0.5	235	5	O44375	O44375 arabia lix	377	7	0.5	265	13	Q98916	Q98916 gallus gall
305	7	0.5	235	5	O44376	O44376 arabia lix	378	7	0.5	266	10	Q9FZ93	Q9fz93 arabisopsis
306	7	0.5	235	5	O44378	O44378 arabia linc	379	7	0.5	266	10	Q9FP61	Q9fp61 oryza sativ
307	7	0.5	235	5	O44379	O44379 arabia linc	380	7	0.5	267	2	Q9HWT0	Q9hwt0 pseudomonas
308	7	0.5	235	5	O46355	O46355 arabia linc	381	7	0.5	267	5	O61076	O61076 plasmodium
309	7	0.5	235	5	O46366	O46366 arabia lix	382	7	0.5	267	10	Q9SJ10	Q9sj10 arabisopsis
310	7	0.5	235	5	O24873	O24873 echinometra	383	7	0.5	267	10	Q9M9U2	Q9m9u2 arabisopsis
311	7	0.5	235	5	O46358	O46358 arabia duf	384	7	0.5	269	2	O20892	O20892 caenorhabdi
										270	2	O30448	O30448 bordetella



93	8	0.6	1082	11	O55201	166	7	0.5	131	10	P93015	P93015 arabidopsis
94	8	0.6	1095	4	Q9HCG6	167	7	0.5	131	10	Q9LKU5	Q9LKU5 arabidopsis
95	8	0.6	1132	4	Q9UF81	168	7	0.5	134	2	Q9RUW2	Q9RUW2 deinococcus
96	8	0.6	1173	11	Q63624	169	7	0.5	135	5	Q9CT85	Q9CT85 plasmodium
97	8	0.6	1191	5	Q9XYV3	170	7	0.5	136	5	Q94171	Q94171 caenorhabdi
98	8	0.6	1196	5	Q23866	171	7	0.5	136	5	Q9NH18	Q9NH18 plasmodium
99	8	0.6	1218	5	Q9V646	172	7	0.5	138	5	Q20867	Q20867 caenorhabdi
100	8	0.6	1220	4	Q9UMN7	173	7	0.5	139	5	O61947	O61947 caenorhabdi
101	8	0.6	1231	5	O18017	174	7	0.5	140	10	Q9SH27	Q9SH27 arabidopsis
102	8	0.6	1235	4	Q9UL54	175	7	0.5	141	9	Q9WCJ2	Q9WCJ2 streptococc
103	8	0.6	1235	11	Q9JLS3	176	7	0.5	141	10	Q9M205	Q9M205 arabidopsis
104	8	0.6	1311	5	Q9VBQ8	177	7	0.5	142	6	Q9TUB9	Q9TUB9 ryptotolagus
105	8	0.6	1311	5	Q9GQN5	178	7	0.5	143	2	Q9RG39	Q9RG39 streptococc
106	8	0.6	1312	4	Q9NR59	179	7	0.5	143	5	Q18533	Q18533 caenorhabdi
107	8	0.6	1331	11	Q9EPQ2	180	7	0.5	144	2	Q9XOW8	Q9XOW8 thermotoga
108	8	0.6	1343	4	Q9H7N4	181	7	0.5	145	4	Q9NTS3	Q9NTS3 homo sapien
109	8	0.6	1398	4	Q9HCF2	182	7	0.5	146	10	Q9FNH0	Q9FNH0 arabidopsis
110	8	0.6	1417	10	Q9FJO9	183	7	0.5	148	4	Q9UJX8	Q9UJX8 homo sapien
111	8	0.6	1520	4	O15087	184	7	0.5	149	2	Q9K9X6	Q9K9X6 bacillus ha
112	8	0.6	1634	5	Q9XW25	185	7	0.5	149	10	Q9FJ67	Q9FJ67 arabidopsis
113	8	0.6	1639	4	Q9H473	186	7	0.5	151	10	Q9SDI1	Q9SDI1 oryza sativ
114	8	0.6	1736	4	O75111	187	7	0.5	154	4	Q9NRN6	Q9NRN6 homo sapien
115	8	0.6	1781	4	Q9UKX0	188	7	0.5	154	5	O44604	O44604 caenorhabdi
116	8	0.6	1857	4	O95153	189	7	0.5	155	10	O65426	O65426 arabidopsis
117	8	0.6	1871	10	Q9SRD5	190	7	0.5	156	10	Q9MSF0	Q9MSF0 medicago sa
118	8	0.6	1890	4	Q9UKW3	191	7	0.5	156	11	Q9JJF0	Q9JJF0 mus musculu
119	8	0.6	2073	4	O9UKW2	192	7	0.5	157	2	Q9ZLD8	Q9ZLD8 helicobacte
120	8	0.6	2248	5	Q9VY17	193	7	0.5	159	5	O9VWNO	O9VWNO drosophila
121	8	0.6	3201	5	O9W0U2	194	7	0.5	159	11	Q9ERD3	Q9ERD3 mus musculu
122	8	0.6	4868	13	Q91319	195	7	0.5	159	14	Q9PX69	Q9PX69 human herpe
123	7	0.5	35	10	O22245	196	7	0.5	160	4	Q9POT1	Q9POT1 homo sapien
124	7	0.5	41	2	O50922	197	7	0.5	160	5	Q9NLE2	Q9NLE2 leishmania
125	7	0.5	43	10	Q9SOC1	198	7	0.5	161	1	O26658	O26658 methanobact
126	7	0.5	43	10	Q9S704	199	7	0.5	162	6	Q9N1M2	Q9N1M2 pan troglod
127	7	0.5	48	5	Q9VWS6	200	7	0.5	162	6	Q9WYQ2	Q9WYQ2 hyllobates l
128	7	0.5	53	4	Q9PLJ2	201	7	0.5	162	10	Q9LFR3	Q9LFR3 arabidopsis
129	7	0.5	61	3	Q91627	202	7	0.5	166	14	Q91BS4	Q91BS4 turkey herp
130	7	0.5	72	4	O15203	203	7	0.5	168	4	Q9H7G1	Q9H7G1 homo sapien
131	7	0.5	72	6	O18889	204	7	0.5	170	3	O13569	O13569 saccharomyc
132	7	0.5	75	10	Q9STQ5	205	7	0.5	171	2	O85021	O85021 photobacter
133	7	0.5	81	10	Q9L2K7	206	7	0.5	172	2	Q9WYU0	Q9WYU0 thermotoga
134	7	0.5	87	11	Q63452	207	7	0.5	172	4	O9P054	Q9P054 homo sapien
135	7	0.5	88	10	Q9A412	208	7	0.5	172	4	Q9NMD7	Q9NMD7 homo sapien
136	7	0.5	91	2	P94908	209	7	0.5	173	2	O68901	O68901 salmonella
137	7	0.5	94	5	Q9NNK8	210	7	0.5	173	2	P75600	P75600 mycoplasma
138	7	0.5	95	6	Q29383	211	7	0.5	173	5	Q27189	Q27189 stylonychia
139	7	0.5	95	10	Q9SCS0	212	7	0.5	173	8	Q9XKN3	Q9XKN3 anthophora
140	7	0.5	97	11	Q9WTY6	213	7	0.5	173	10	O64930	O64930 pseudotsuga
141	7	0.5	97	14	O64890	214	7	0.5	175	10	Q9ZT86	Q9ZT86 brassica na
142	7	0.5	98	4	Q9NNW1	215	7	0.5	177	1	Q9FFZ0	Q9FFZ0 aeropyrum p
143	7	0.5	98	6	Q9N106	216	7	0.5	177	3	Q9P542	Q9P542 neurospora
144	7	0.5	103	4	Q9NNZ6	217	7	0.5	178	2	Q9KFN6	Q9KFN6 bacillus ha
145	7	0.5	107	4	Q15254	218	7	0.5	178	10	O49595	O49595 arabidopsis
146	7	0.5	107	10	O9LQF6	219	7	0.5	178	10	O42461	O42461 canavalia g
147	7	0.5	108	2	Q9R465	220	7	0.5	178	10	O42492	O42492 canavalia g
148	7	0.5	108	4	Q15202	221	7	0.5	179	5	Q9VC01	Q9VC01 drosophila
149	7	0.5	109	1	Q9RE13	222	7	0.5	181	5	O02040	O02040 caenorhabdi
150	7	0.5	109	4	Q15249	223	7	0.5	181	6	Q9TRZ5	Q9TRZ5 ryptotolagus
151	7	0.5	109	4	Q15204	224	7	0.5	183	2	Q9K6F2	Q9K6F2 bacillus ha
152	7	0.5	109	4	Q15200	225	7	0.5	184	10	O22801	O22801 arabidopsis
153	7	0.5	109	6	Q29286	226	7	0.5	190	4	Q9H2P1	Q9H2P1 homo sapien
154	7	0.5	110	10	Q9STQ8	227	7	0.5	191	2	Q9ZEB4	Q9ZEB4 rickettsia
155	7	0.5	111	5	Q9U553	228	7	0.5	191	4	O9UEL2	O9UEL2 homo sapien
156	7	0.5	112	10	O40726	229	7	0.5	192	5	Q9GUA2	Q9GUA2 riftia pach
157	7	0.5	112	10	Q9FVR1	230	7	0.5	193	2	P94221	P94221 borrelia bu
158	7	0.5	118	10	O65442	231	7	0.5	193	2	P94227	P94227 borrelia bu
159	7	0.5	118	10	Q9FPL5	232	7	0.5	194	2	Q9HVC3	Q9HVC3 pseudomonas
160	7	0.5	118	14	Q9DUD4	233	7	0.5	196	2	Q9KQV4	Q9KQV4 vibrio chol
161	7	0.5	121	4	Q9R306	234	7	0.5	196	14	O55275	O55275 fowl adenov
162	7	0.5	125	4	Q9H014	235	7	0.5	197	10	Q39819	Q39819 glycine max
163	7	0.5	129	10	Q9LFR4	236	7	0.5	197	11	Q9ERC3	Q9ERC3 mus musculu
164	7	0.5	130	2	O50411	237	7	0.5	198	14	Q9Q829	Q9Q829 Shope fibro
165	7	0.5	131	2	O05787	238	7	0.5	198	14	Q91GT8	Q91GT8 porcine ade



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:04:49 : Search time 40.2 seconds  
(without alignments)  
4578.017 Million cell updates/sec

Title: US-09-512-581-2  
Perfect score: 1391  
Sequence: 1 MAHSKTRNDGKITYPGVK.....QKGRGRSPKTPSPQPKKNV 1391

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL\_16.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1015	73.0	1416	4	Q9V2I5
2	1015	73.0	1447	4	Q9NTI5
3	353	25.4	1391	4	Q9Y451
4	125	9.0	125	4	Q9H5N8
5	19	1.4	851	4	Q9Y4D4
6	9	0.6	419	4	Q93029
7	9	0.6	554	10	Q9SG14
8	9	0.6	837	3	Q9P607
9	8	0.6	59	11	Q9R1N5
10	8	0.6	66	11	Q9WVP2
11	8	0.6	81	10	Q9M3G5
12	8	0.6	99	5	Q23793
13	8	0.6	99	5	Q23794
14	8	0.6	131	10	Q96475
15	8	0.6	147	14	Q84290
16	8	0.6	149	10	Q9LV56
17	8	0.6	197	5	Q23872
18	8	0.6	204	10	Q9SB54
19	8	0.6	231	3	Q9P529

20	8	0.6	235	10	Q9SXG4
21	8	0.6	251	4	Q95472
22	8	0.6	294	10	Q22924
23	8	0.6	299	10	Q9M4F1
24	8	0.6	310	10	Q39175
25	8	0.6	313	10	Q39089
26	8	0.6	316	2	Q50887
27	8	0.6	316	10	Q9LI19
28	8	0.6	325	5	Q44016
29	8	0.6	329	10	Q9FM77
30	8	0.6	331	5	Q9VD43
31	8	0.6	341	1	Q9YFQ6
32	8	0.6	351	5	Q9G263
33	8	0.6	368	3	Q9P837
34	8	0.6	379	10	Q9XE25
35	8	0.6	388	6	Q9GL18
36	8	0.6	389	6	Q9GK25
37	8	0.6	395	10	Q9LTR6
38	8	0.6	425	4	Q9NQD6
39	8	0.6	442	2	Q34631
40	8	0.6	450	11	Q9WVR7
41	8	0.6	462	11	Q9R1B7
42	8	0.6	466	3	Q9HEL6
43	8	0.6	468	1	Q57853
44	8	0.6	470	11	Q9QYK4
45	8	0.6	475	4	Q43818
46	8	0.6	501	10	Q39482
47	8	0.6	505	11	Q37770
48	8	0.6	507	10	Q9SL49
49	8	0.6	508	5	Q9VS02
50	8	0.6	510	4	Q75239
51	8	0.6	519	4	Q95676
52	8	0.6	521	11	Q08795
53	8	0.6	522	5	Q61090
54	8	0.6	531	4	Q9H166
55	8	0.6	535	11	Q9JY99
56	8	0.6	545	4	Q9H117
57	8	0.6	552	10	Q49397
58	8	0.6	562	3	Q9P557
59	8	0.6	564	10	Q43673
60	8	0.6	566	2	Q9WZU1
61	8	0.6	568	1	Q28851
62	8	0.6	568	2	Q9USE4
63	8	0.6	583	11	P70616
64	8	0.6	587	10	Q9LWV1
65	8	0.6	587	11	P70615
66	8	0.6	588	11	Q61791
67	8	0.6	602	14	Q56852
68	8	0.6	605	4	Q9NZ53
69	8	0.6	612	13	Q919E7
70	8	0.6	644	4	Q9H3E6
71	8	0.6	646	10	Q64877
72	8	0.6	655	2	Q9KQ4
73	8	0.6	662	5	Q9VFD5
74	8	0.6	665	4	Q76034
75	8	0.6	666	10	Q9FK63
76	8	0.6	671	3	Q12242
77	8	0.6	711	4	Q9U2H3
78	8	0.6	720	5	Q9U2H3
79	8	0.6	730	10	Q82332
80	8	0.6	765	10	Q9FZA9
81	8	0.6	787	10	Q9M0L6
82	8	0.6	791	13	Q9DGL1
83	8	0.6	802	10	Q38898
84	8	0.6	840	13	Q918N3
85	8	0.6	841	1	Q59612
86	8	0.6	874	11	Q88453
87	8	0.6	938	4	Q13187
88	8	0.6	954	4	Q15326
89	8	0.6	973	4	Q9H0Z1
90	8	0.6	1037	10	Q9FH58
91	8	0.6	1047	3	Q9P6E3
92	8	0.6	1062	4	Q9NSW2

Q9SXG4 oryza sativ  
Q95472 homo sapien  
Q22924 arabidopsis  
Q9M4F1 brassica na  
Q39175 arabidopsis  
Q39089 arabidopsis  
Q50887 myxococcus  
Q9LI19 oryza sativ  
Q44016 dictyosteli  
Q9FM77 arabidopsis  
Q9VD43 drosophila  
Q9YFQ6 aeropyrum p  
Q9G263 plasmodium  
Q9P837 candida alb  
Q9XE25 oryza sativ  
Q9GL18 diceros bic  
Q9GK25 tupia tana  
Q9LTR6 arabidopsis  
Q9NQD6 homo sapien  
Q34631 bacillus su  
Q9WVR7 rattus norv  
Q9R1B7 rattus norv  
Q9HEL6 neurospora  
Q57853 pyrococcus  
Q9YK4 mus musculu  
Q43818 homo sapien  
Q39482 calocedrus  
P37770 mus musculu  
Q9SL49 arabidopsis  
Q9VS02 drosophila  
Q75239 homo sapien  
Q95676 homo sapien  
Q08795 mus musculu  
Q61090 plasmodium  
Q9H166 homo sapien  
Q9JY99 mus musculu  
Q9H117 homo sapien  
Q49397 arabidopsis  
Q9P557 neurospora  
Q43673 vicia faba  
Q9WZU1 thermotoga  
Q9Z851 chlamydia p  
Q9JSE4 chlamydia p  
P70616 rattus norv  
Q9LWV1 oryza sativ  
P70615 rattus norv  
Q61791 mus musculu  
Q56852 fowl adenov  
Q9NZ53 homo sapien  
Q919E7 brachydanio  
Q9H3E6 homo sapien  
Q64877 arabidopsis  
Q9KQ4 vibrio chol  
Q9VFD5 drosophila  
Q76034 homo sapien  
Q9FK63 arabidopsis  
Q12242 saccharomyc  
Q9H16 homo sapien  
Q9U2H3 caenorhabdi  
Q82332 arabidopsis  
Q9FZA9 arabidopsis  
Q9M0L6 arabidopsis  
Q9DGL1 fuqu rubrip  
Q38898 arabidopsis  
Q918N3 rana pipien  
Q59612 pyrococcus  
Q88453 rattus norv  
Q13187 homo sapien  
Q15326 homo sapien  
Q9H0Z1 homo sapien  
Q9FH58 arabidopsis  
Q9P6E3 neurospora  
Q9NSW2 homo sapien



[illegible]



**THIS PAGE BLANK (USPTO)**



```

334  ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
      |||||::: :::: |||||
124094  ....TGCCTTTTGTGAACAAAAAC.....AAGAA 124069

350  pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367
      :||| :::: :::: :::: ::::
124068  CCTGATG...TTTCTACTGATGCGG..... 124047

367  rHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeu 383
      :||| :::: :::: :::: ::::
124046  ....AATCAATCAACATTCCATCGAAAGTTCCGCTC 124014

384  LeuValAsnAspHisLeuLeuTcnsnPhenPheValArgGluArgThr 397
      :||| :::: :::: :::: ::::
124013  TCAACGAACATTCAGAGAATATCCAAAGTGGAGGAGTACACACTTTAGAT 123964

398  ....LeuAspLysArg.....TrpArgVal...ArgL 406
      |||||::: |||||::: |||||::: |||||
123963  AATACTTAACAAGATAAAACGGTATCCCGTCATCGGAAGTTTAAAGGA 123914

406  ysGluAlaMetMetGlyLeuAla..... 413
      ||::: ::||::: ||:::
123913  AACAAAACTGATGATCTATCCAAATCCGTCACCAATACGTATCCAAACCG 123864

414  ...GlnIleTyrLysLysTyrAlaLeuGlnSerAla..... 424
      :||| :::: |||||::: |||||::: |||||
123863  AGGAACCTTCAAGAAGAAGCTAAAGCTGATGAGTTGCTACTGTAGTGAA 123814

425  ....AlaGlyLysAspAlaAlaLysGlnIleAlaTyrPileLysAspL 439
      :||| :::: |||||::: |||||::: |||||
123813  ATGGTTCAAGCATCTCAAGAGACGACGAAACAGATGACTATTAAACCAAT 123764

439  ysLeuLeuHisIleTyrTyrGlnAsnSer.....Ile 449
      :||| :::: |||||::: |||||::: |||||
123763  CATCGAAACATCCCTGGAAGAGACCCCTTAGAAGAGGCACATCCTGATG 123714

450  AspAspArgLeuLeuValGluArgIlePhe..... 459
      :||| :::: |||||::: |||||::: |||||
123713  AGTTGCTCTCTTCTAGTGAAGAGGATCAAGTAATTTGCGAAACACCATCT 123664

460  ....AlaGlnTyrMetValProHisAsnLeuGluT 470
      :||| :::: |||||::: |||||::: |||||
123663  GCTGAACATAGGCCAACCATCGAAACATCTCTGTGACGAAACCATATCAAC 123614

470  hrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
      :||| :::: |||||::: |||||::: |||||
123613  AGCAGCAACTGATATAGTCCTGATT..... 123588

487  AsnAlaValLysAlaLeuAsn.....G 494
      |||||::: |||||::: |||||::: |||||
123587  ...TAAGTCAAAATTGATCAACTAAATTTTGAATAATCACCGCCGAAGAA 123541

494  uMetTyrLysCysGlnAsnLeuLeuArg..... 503
      :||| :::: |||||::: |||||::: |||||
123540  CCGCATGAGAAACCAAACTGATGAGACTACCCAGGAAGTCAATCGAAAC 123491

504  ..His..GlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrA 519
      :||| :::: |||||::: |||||::: |||||
123490  ATCACACGGCGGAGAAAGATCAATAGAA.....GCCAGTCTTAAGAGG 123447

519  spLasSerValLysAlaIlePheSerLysValMetValIle..... 532
      :||| :::: |||||::: |||||::: |||||
123446  ATTCACTGTGTAGTAATAATGGATAATGCAGTTGTGTGGAATACTAATCTCAA 123397

533  ThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPh 549
      :||| :::: |||||::: |||||::: |||||
123396  ACAGATAAGCTAATCTCCGCGATGAATACAAATTCGCGGCGCG..... 123352

549  eThrGlnValLeuGluAspAspGluLysIleArgLysGln.....L 563
      :||| :::: |||||::: |||||::: |||||
123351  .GPAGATATATGATCGATGAGTCTCTTTAAGAAATTTCTAGTTACAGCAT 123303

```

```

563  euGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCys 579
      :|||::: |||||::: |||||::: |||||
123302  TTTCAGATCGAACACATCAGCAAGAAGTCGCGCAGCACTTCTTTAGACTTAGAT 123253

580  ValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPr 596
      :|||::: |||||::: |||||::: |||||
123252  CTTTCCGAGATAAGTAAGAAAGGAATAATGGAAGCTTTGCCGAAT.. 123205

596  oPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisI 613
      :|||::: |||||::: |||||::: |||||
123204  ....GACAACGCTTGTGTAAACACA 123183

613  le.....AspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627
      :|||::: |||||::: |||||::: |||||
123182  TAGCGTGCATCTTCTAATGTAAGCTCATTAAGAACCAACGAATTTA 123133

628  Ser.....IleAspGlyThrAlaAspAspGluAspGluGlyVa 640
      :|||::: |||||::: |||||::: |||||
123132  TCTCCCAAGACTCGTCTAGCTGGACATCAATGTCACAAGATGAGGCCGC 123083

640  lProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysValLeus 657
      :|||::: |||||::: |||||::: |||||
123082  TCAAGTTGATGAACAG.....A 123066

657  erPheThrHisProIleSerPheHisSerAla...GluThrPheGluSer 672
      :|||::: |||||::: |||||::: |||||
123065  ATCAAAACGATCAGCTCGAAGAACACGTCGCAATGGAGGTATCATCAAT 123016

673  LeuLeuAlaCysLeuLysMetAspAspGluLys.....ValAlaGl 686
      :|||::: |||||::: |||||::: |||||
123015  GAAATCGCTGCATAAAGAGAGTGTAGAAATTTGAAGACATAGATGA 122966

686  uAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAsp 703
      :|||::: |||||::: |||||::: |||||
122965  GCATGCTTGGCGGAGTTAAACAA.....AGTTCACCTGGAAGAAAT.. 122923

703  heProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLys 719
      :|||::: |||||::: |||||::: |||||
122922  ....GCATTAATTCATCAATCTTA 122902

720  LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI 736
      :|||::: |||||::: |||||::: |||||
122901  ACGGACCAGCAATAAGAGGTTTCTATTATGGAGAAGATCCGAAGCTGTTGT 122852

736  ePheSerSer...LysGluThrGlnPheAlaGlnIlePheGluProLeuH 752
      :|||::: |||||::: |||||::: |||||
122851  TTTAGAGGAACCTAAGACAGCAACCAAAATCAGGATTTCAAAGACAGAG 122802

752  isLysSerLeuAspProSerAsnLeu.....GluHisLeuIleThrPro 766
      :|||::: |||||::: |||||::: |||||
122801  TGATTCGCTTAGACGCTCCATCATTTGTTGAGAAAAGATACACTAGATGAA 122752

767  LeuValThrIleGly.....HisIleAlaLeuLe 776
      :|||::: |||||::: |||||::: |||||
122751  AGTACATCAGTTGGAGAGGATAAGATCAATCCACAGAANTTCTATGCT 122702

776  uAlaProAspGlnPheAlaIaProTyrPlySerTrpValAlaIaThrPheI 793
      :|||::: |||||::: |||||::: |||||
122701  TAAAGAACAACTAATTCGCGACCAA..... 122677

793  leValLysAspLeuLeuMetAsnAspArg.....Leu 803
      :|||::: |||||::: |||||::: |||||
122676  ..ATAAAGATGTTAGCTCGATGGCAAAAAAGAAATTCACAGAATGTATA 122629

804  ProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPr 820
      :|||::: |||||::: |||||::: |||||
122628  CCAAAAAAGGATTCT.....GACGAAAAAGCAACATGA 122597

820  oGluThrMetValLysIleGlnAla..... 828
      :|||::: |||||::: |||||::: |||||
122596  CGATAAAGTTATTCAATTAGATGCTCGAGCTTAGATGAAAGTTACACAG 122547

829  ....IleLysMetMetValArg 834

```



1363 AllLeuSerThrGlnSerThrProGlnLysGlyArgGlyArgProSe 1379  
 ...|||:||||:..:||||:||||:||||:||||:||||:||||:  
 5851 ACA...GAATAACAAATATCAACGCCGCAGTAATGCTTTAGACACCGTACC 5897  
 1379 rLyThr 1381  
 |||||  
 5898 AAAACA 5904

seq\_name: gb\_in1:AC011615

seq_documentation_block:			
LOCUS	AC011615	199653 bp	DNA INV 17-FEB-2001
DEFINITION	Drosophila melanogaster, chromosome 3R, region 89D-89D, BAC clone BACR01H23, complete sequence.		

ACCESSION	AC011615
VERSION	AC011615.4
	GI:12957619

KEYWORDS	HTG.
SOURCE	fruit fly.

ORGANISM *Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

**AUTHORS**

Celniker, S. E., Adams, M. D., Krommiller, B., Tyler, D., Wan, K. H., Holt, R. A., Evans, C., Gocayne, J. D., Ananides, P. G., Brandon, R. C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K. Y., Busam, D. A., Carlson, J. W., Center, A., Champe, M., Davenport, L. B., Dietz, S. M., Dodson, K., Dorsett, V., Doup, L. E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R. F., Garg, N. S., George, R. A., Gonzalez, M., Houck, J., Hoskins, R. A., Hostin, D., Howland, T. J., Iqbalwala, C., Jaitani, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A., McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouenavong, S., Pittman, G. S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S. M., Zaveri, J. S., Smith, H. O., Rubin, G. M., and Venter, J. C.

TITLE

**JOURNAL**  
**Unpublished**

REFERENCE 2 (bases 1

**AUTHORS**

Celniker, S. E., Agbayani, A., Arcaina, T. T., Baxter, E., Blazej, R. G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C. M., Fafan, D. E., Galle, R., George, R. A., Harris, N. L., Hoskins, R. A., Houston, K. A., Hummasti, S. R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M. A., Mazda, P., Moshrefi, A. R., Moshrefi, M., Nixon, K., Pacleb, J. M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snit, E., Svirskaas, R. R., Wan, K. H., Weinburg, T., Zhang, R., Zierian, L. L. and Rubin, G. M.

TITLE  
RUBIN, G.M.  
Direct Submission

**JOURNAL.** Submitted (08-OCT-1999) Drosophila Genome Center, Lawrence Berkeley  
**FILE.** Direct submission  
**FILE.** Laboratory, MS 64-121, Berkeley, CA 94720, USA

LABORATORY, MD 04 121, BELKLEY, CH 34720, USA  
On Feb 17, 2001 this sequence version replaced qi:6838819.  
COMMENT

COMMENTS: ON FEB 17, 2001 THIS SEQUENCE SUBMITTED BY:

Sequences submitted by:  
**Berkeley Drosophila Genome Project**

Lawrence Berkeley National Laboratory

Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdqpe@fruitfly.berkeley.edu](mailto:bdqpe@fruitfly.berkeley.edu).

## FEATURES

source

/organism="Drosophila melanogaster"

```
/strain="y; cn bw sp"
```

/db\_xref="taxon:7227"

/chromosome="3R"

```
/map="89D-89D"
```

```
/clone="BACR01H23 (D1149)"
```

```
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
```

*Drosophila melanogaster* BAC library, partial EcoRI in-

pBACE3.6)<sup>n</sup>

```

BASE COUNT      55654 a 43133 c 44105 g 56761 t
ORIGIN

alignment_scores:
    Quality: 224.50      Length: 1540
    Ratio: 0.329        Gaps: 76
    Percent Similarity: 44.351      Percent Identity: 19.416

alignment_block:
US-09-512-581-2 x AC011615/rev ..

Align seg 1/1 to reverse of: AC011615 from: 1 to: 199653

147  LeuGluAspSerAsnGluIlePheThrGlnLeuTyArgThrLeuPheSe 163
    ||| :|||::: :|||::: :|||::: ||| :|||
124787 CTAATGCTCTCGACAGTAGTGTGGAANAACATATACCTGACCGAATCTC 124738
    rValIleAsnAsnGlyHisAsnGlnLysValHisMet...HisMetValA 179
    :||| :||| :|||:|||||::: :|||:::
124737 AAAGGTGGCAAACTTCAAGAAGAGAGAAAAGTTAGATATCGCATCAGTTAA 124688
    179 splLeuMetSerSerIle.IleCysGluGlyAsp_ThrVal..... 191
    |||::: :|||::: |||::: |||::: |||:::
124687 TTCAGATGAAATATTGAATCTCGAGTGGAAATTGACCATCTCAGAGTCAG 124638
    192 ..... SerGlnGlu 194
    ||| :||| :|||
124637 AGATATTTCTTTAGACACAAACAGTGGATAGTTCGATCAATTCTACTAAT 124588
    195 LeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
    :|||::: :|||::: :|||::: ||| :|||
124587 GTCCTTGACAGTAGTGTGGAANAACATATACCTGAACAT...AATCTTAA 124541
    211 nLysGlnAlaTyArgLeuAlaLysAlaLeuLeuLysArgThrAlaGlnA 228
    :|||::: :|||::: :|||::: ||| :|||
124540 TCAAGAAGAT.....GCCAAGGCCACAAACGATG 124512
    228 latLeuGluProTyIleThrPhePheAsnGlnValLeuMetLeuGly 244
    ||| :||| :|||
124511 CCCAAGACACTTCAGTTACA.....GAT 124489
    245 LysThrSerIleSerAspLeuSerGluHisValPheAspLeuIle..... 259
    :||| :||| :||| :||| :||| :|||
124488 GAAAAATCCTCTACTGACTTATATCGTTCAGTTGGCGAAGTATTACCATC 124439
    260 ..... LeuGluLeuTyArgAsnIleAspSerHisLeuLeuLeuSerVal 274
    :|||::: :|||::: :|||::: ||| :|||
124438 CAATAAAGTGGATTTCGATCTACTGATGACCAACCAATCGTAAATATCTCT 124389
    274 euProGlnLeuGluPheLysLeuLysSer..... 283
    :||| :||| :|||
124388 CTAAAGACTGGCAAAAGCGGCAAAATCTGGTGTGTAACCCCATCGAC 124339
    284 AsnAspAsn.....GluGluArgLeuGlnValVa 293
    ||||| ||| :|||::: :|||:::
124338 ACGCAACTCCAAACCATATATTGACGCTCTCAAGAAGTTCGCGAGA 124289
    293 lLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSer 309
    |||::: :|||::: :|||::: ||| :|||
124288 AACCAACACAGATAAGGTTTCGAGTGTGCAAGATCGATCGACAGTCA 124239
    310 Gln.....AsnLysProLeuTrpGlnCysTyLeuGlyAr 321
    ||| :||| :|||::: ||| :|||
124238 CAGAAACTTCGCCAGTCCAAAAGCCCTCGCGCAGAGTCTTAAGCAACCG 124189
    321 g..... 321
    ||| :||| :|||
124188 TTCGACTCAATCCGATGAACCTGTAAGAAACCAACTCAATGA 124139
    322 .....PheAsnAspIleHisValProIleArgLeuGluCys 333
    ||||| :|||::: :|||::: ||| :|||
124138 GGTTCCTGGTATGGGAACATCTATTCAACCAAGCATTCAGACTT..... 124095
    :|||::: :|||::: :|||::: ||| :|||

```







```

439 ysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuVal 455
||||| : : : : : |||||
2624 AATCATGAAAGTTGTGCTCCAG .....GCTGTT 2652
456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrG1 472
||| : : : : : : : : : |||||
2653 GAAGTAGCCATTTCTAAAGTTAATTTGGCTCATGATCTGATCTTACATC 2702
472 uArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlav 489
2703 T..... 2703
489 allYsAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGln 505
2704 .....TTGAAGCACTT 2715
506 ValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
: : : : : ||| : : : : :
2716 CCAGAGGAT.....AGAAAGGACCAAACTCATTT 2744
522 LysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAsp 539
: : : : : : : : : : : : : : : |||
2745 AGAATCCTTTAGCAGCTTATATAAAGTCACGCAAAACAACAAAGATA 2794
539 roGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu..... 553
2795 CACCATTTTTCGTGACTTTTGGTGCTGTACTTCATCTGTTGTTTAT 2844
554 .....GluAspAspGluLysIleArgLy 561
: : : : : : : : : : : : : : :
2845 ACATTTTGGACCTGTTTTAAGAAAGAAATCATCATCTAGTAAATTT 2894
561 sGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluG 578
2895 CAATATCCAAATTTCCAAAGTTGTTATTCGCAAGGAAACATGACCTTG 2944
578 lYcysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThr 594
: : : : : : : : : : : : : : : |||
2945 ATCTTTTAAAGAGTTTATCAGGAAGCTCGAAC.....GAACA 2985
595 AsnProPheLeuGluMetIleLysPheLeuLeu..... 605
: : : : : : : : : : : : : : :
2986 TCTGAATTTCTGATAATAGAAAAGTTCTAGAAATGGATGATAGAAAAC 3035
606 .....GluArgIleAlaProValHisI 613
3036 AGAAGTATATGCCAAATTTGGTGGATCAACGTTATTACCAACAATA 3085
613 leAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIle 629
||| : : : : : : : : : : : : : : : |||||
3086 TATCTTTCAGAGATTTTCAGAGTTTGGCGAATATTTGGAACAA..... 3129
630 AspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaI 646
||| : : : : : : : : : : : : : : : |||||
3130 .....GTTCCGAATGAAATTCAT 3149
646 eArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleS 663
||| : : : : : ||| : : : : :
3150 TGAATAATTTCTGGACCTATGTTTGAAGCTCAGCTTT.....CCAGTTA 3193
663 erPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMet 679
: : : : : ||| : : : : :
3194 ATTTG.....TTTACACTACTTCATGTCTATGTGTCG 3228
680 AspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrG1 696
: : : : : : : : : : : : : : :
3229 AACAAATATTTTATTTATTTTCAAGAGCTATGTTTCAAAAAATGAA.. 3276
696 ySerLysIleGluGluAspPhe..... 703
: : : : : ||| : : : : :
3277 AACAAATTAACCTGGATCTGATCAGCTGTTAAACACTTCTTTACCAG 3325
704 ....ProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSer 718

```

```

3326 GGAATCCTGAAGCTGTTTTCGGTCTGCTCCCATTTTA..... 3363
719 LysLysGlyProProArgGlnAlaLysTyr.....AlalleHisCy 732
||| : : : : : ||| : : : : :
3364 .....AGCGTAACAAATTTATGATATTTTGGAGTACTG 3398
732 sIleHisAla..... 735
||||| : : : : :
3399 TATTCATCCAAATTCCTAAATTTTACTAAATTCCTATCCCTGATCTAAATTCAG 3448
735 ..... 735
3449 ATCTCTATTGAAGCTGTTACCAGTAGCAGAGCTTCATATTTTGCAGCA 3498
736 .....IlePheSerSerLysGlu..... 741
: : : : : : : : : : : : : : :
3499 AACATAAACTATTCAAGTGTCTCGAACAGCTTACATTAGTTCTGTGGCT 3548
742 .....ThrGlnPheAlaGlnIleP 748
3549 GTTGAAGGTCACAACTTGAACAGCTAAATCAAAATTTTTCAGAAATCG 3598
748 heGluProLeuHisLysSerLeuAspProSerAsnLeuGlu.....His 762
: : : : : : : : : : : : : : :
3599 AGAACGTTTACAAATGCGTCTGAT...AGCGAGTTAGAAAAATCAGAA 3645
763 LeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAs 779
||| : : : : : ||| : : : : :
3646 ATAATCAGAGAAGCTCCTT.....CATCTGGCGATGGCAAAATCCAT 3686
779 pGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysA 796
: : : : : ||| : : : : : : : : :
3687 TGAA.....CCATTATTCAGTGGCTTATTGAATTTTTCATAAGA 3727
796 spLeuLeuMetAsn.....AspArgLeuProGlyLysLysThrLys 810
: : : : : : : : : : : : : : :
3728 ATAATATGCGGATCATTTGGATGAATTTTCGGAATATGACCAGCGAA 3777
811 LeuTrpValProAspGluGluValSerProGluThrMetValLysIleG1 827
||| : : : : : ||| : : : : :
3778 GTTCTTTTC.....AAAATAAGTCCAGAGCTACTGTATAAATTAAT 3818
827 nAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 844
: : : : : ||| : : : : :
3819 GAGTATAG.....GAGA 3832
844 erLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
: : : : : : : : : : : : : : :
3833 AGCCTAATGGAGCTTACTGCAGCAGTTATTGAAAAAATTTGAAAAATGGA 3882
861 AspGlyAsp.....LeuThrGluGlnGlyLysIleSerLysPr 873
||| : : : : : ||| : : : : :
3883 GATGATGATTATATATTGGAGTTGCTTGAAGAAGATTATTATACAAAAAGA 3932
873 oAspMetSerArgLeuArgLeuAlaGlySerAlaIleValLysLeuA 890
||||| : : : : :
3933 G.....ATCCAAATTTTGAAAAAT 3952
890 laGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeu 906
: : : : : ||| : : : : :
3953 TAAAGGAGCCA.....CTTTAGTTTCTCTCTTA 3981
907 CysAlaLeuAlaIleAsnAspGluCysTyrGln.....Va 918
: : : : : ||| : : : : :
3982 AATCCAGTTAGCTCTAATATGCAAAAAACAAAAAGAGTACAAATATGCT 4031
918 lArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgL 935
: : : : : : : : : : : : : : :
4032 CCGTGAATTAGTCTTATATATATCTCACTAAACCGCTTCTAGAAAGTCCG 4081
935 euProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro... 950
: : : : : : : : : : : : : : :

```



```

alignment_block:
US-09-512-581-2 x SCRIF1
Align seg 1/1 to: SCRIF1 from: 1 to: 6541

3 HisSerLysThrArgThrAsnAspGlyLysIleThrTyr..... 15
:::|||||:::|||||:::|||||
1000 TATAAGGAACAAGGTAATCCTCGCGCTTAGAGATATTGGGTGAGTGG 1049
16 .....ProProGlyValLysGluIleSerAspL 25
:::|||||:::|||||:::|||||
1050 TGAGATCCATGCTTGTAGTGAATAAGAAAGTGTTCGGAGTTTAAGAAA 1099
25 ysile.....SerLysGluGluMet 31
||| |||::: |||
1100 TTATTGAGGCGGATTTGGSCATCTTGGACACAAGATCTGAAGATTACGTC 1149
32 ValArgArgLeuLysMetVal.....Vally 40
:::|||||:::|||||:::|||||
1150 GCGAGGAGGTTTGGAGGTATACGAACCTTTTAATAATATTATACCCATTCT 1199
40 sThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrL 57
||| |||::: |||
1200 ACCCACTAAAAATGTTAACGAGGTTGATCAGAAATTCAAATATCTAATG 1249
57 euAsnLeuAla.....LeuHisLeuAla 64
:::|||||:::|||||:::|||||
1250 TCAATATAGAAAAGTATCATTTGAATATGCTATACCGCATCTACAAATAGCA 1299
65 SerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 81
:::|||||:::|||||:::|||||
1300 CAGGACACATTAATTAAGTAGTAGTGAAGAAAAAATCCATTGTGAATC... 1347
81 acYscYsLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrT 98
||| |||::: |||
1348 .....AGACTATATGTTCAATAGTTCGATTCT 1375
98 hrSer..... 99
|||
1376 TCAGTGCTATTATGCTAATTTCAAAATCGTTAAGTGGCTAACGAAAAGG 1425
100 ProAsp.....LysLeuLysAspIlePhe..... 107
||| ||| ||| ||| ||| |||
1426 CCAGATTGTTGTTAAAGTTAAAGTTAATTTACAGGTGGCACCACCGGTGC 1475
108 .....MetPheIleThrArgGlnLeuLysG 116
||| ||| ||| ||| ||| |||
1476 GCTAGAATGAATTCACAAAGATATATCATACAGCACAGTCTCTT 1525
116 lYLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
||| ||| ||| ||| ||| |||
1526 TCTTGAGAGACCAAAA.....TTTGGTACTTTTCTTGGAGCAAC 1566
133 GluAsnIle...AlaTrpValLysSerTyrAsnIleCysPheGluLeuG 148
||| ||| ||| ||| ||| |||
1567 GAAGAAATTTAAACCTCATCTATAGCACTTTCACTGAAATAATAGGAATCAA 1616
148 uAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerVal 165
::: ||| ||| ||| ||| ||| |||
1617 TAGTCACACCTTAATATATGAATAATG.....CTAC 1648
165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181
||| ||| ||| ||| ||| |||
1649 TTATAAGGGGATTTTGGACCAAGTATCCAAAATTAATGATTGAGACGGTA 1698
182 SerSer.....IleIleCysGluGlyAs 189
::: ||| ||| ||| ||| ||| |||
1699 ACTTCTTGTTTACCAGGTGAGGTACTACTAGGATTTATATTCGGAGATGA 1748
189 pThrValSerGlnLeuLeuAspThrValLeuValAsnLeuValProA 206
||| ||| ||| ||| ||| |||
1749 AATTACTCCATGAATTTCTCATACTTCAATAGTTGTTTACTAGT.... 1794

```

```

206 laHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeu 222
1795 .....GAACTACTA 1803
223 LysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnG 239
||| ||| ||| ||| ||| |||
1804 AAAAATGCTTAGATTTTGTGATGAACATGAAGGATTAC.....CA 1847
239 nValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValP 256
||| ||| ||| ||| ||| |||
1848 GTGTATCATGCTATCTCCAGTATGCGAAACAATCCCGGAAAAATTTTAT 1897
256 heAspLeuIleLeuGluLeuTyr.....AsnIleAsp..... 266
||| ||| ||| ||| ||| |||
1898 CTAACACTACCGTTAAATTCATATGACAGCGCTAAATTAGATAAAGTGACA 1947
267 ....SerHisLeuLeuSerValLeuProGlnLeuGluPheLysLeuL 282
||| ||| ||| ||| ||| |||
1948 ATGTGTCATCTCTTACACAGCAGATAAAA.....AATATATTTGTTGT 1991
282 sSerAsnAspAsnGluGluArgLeuGlnVal...ValLysLeuLeuAla 298
||| ||| ||| ||| ||| |||
1992 CAAAATGACATAAATAAGCAATGATTTGTGGCTTTCATGACCGGAT 2041
298 ysmetPhe.....GlyAlaLysAspSerGluLeuAlaSerGlnAsnLys 312
||| ||| ||| ||| ||| |||
2042 TACTATACGATAGCGAAAGAGAGATTTATGATTGACATCAGAAATCTAAC 2091
313 ProLeuTrp.....GlnCysTyrLeuGlyArgPheAsnAs 324
||| ||| ||| ||| ||| |||
2092 AAAGTTTGGTTTGATTTAAATAACCTGTCTTTATCAATAAT..... 2133
324 pIleHisValProIleArgLeuGluCysValLys.....PheAlaAs 338
||| ||| ||| ||| ||| |||
2134 ....CACCCCAAGACTCGCTTAATGTCCATCAAGTTTGGAGAAATTA 2179
338 erHisCysLeu.....MetAsnHisProAspLeuAlaLys 349
||| ||| ||| ||| ||| |||
2180 CCTACTGCATATCCACGAAATATCACAGAAAAATCAGGAAGGAACAAA 2229
350 AspLeuThrGluTyrLeuLys.....Va 357
||| ||| ||| ||| ||| |||
2230 AGTTTACTGTCTACTTTTACGAACCCCTTTTCAAATGACGCTTCTTATGT 2279
357 lArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 374
||| ||| ||| ||| ||| |||
2280 GAATGATCCAAGCGCTAGAGAAGGATCATCTACCACCTTTTAGGTGTG 2329
374 leValThrAla.....AlaLysLysAspIleLeu..... 383
||| ||| ||| ||| ||| |||
2330 TTTACACTGCATTTTACAAGTAATAAGAATTTGCAACAGATATGTTGAA 2379
384 LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAsp 400
||| ||| ||| ||| ||| |||
2380 CTATTTTGGGACCACCTAATAACCCCGATTTATGAAGATTATGTTTCAA 2429
400 sArgTrpArgValArg..... 405
||| ||| ||| ||| ||| |||
2430 ATATGATTCGATTTCACAAAATGTACTCTTCACAGTTTTCATTGCT 2479
406 .....LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
||| ||| ||| ||| ||| |||
2480 TAATTTGGGGAAGAATGACGATGTTGCAATGGAAAGAAATATAAAAG 2529
419 Tyr.....AlaLeuGlnSerAlaAlaGlyLysAs 428
||| ||| ||| ||| ||| |||
2530 CATATTCACCCCAATGAGTGTATATAGCCTCAGAAAGGTGTAAGCTAAAAG 2579
428 pAlaAlaLys.....GlnIleAlaTrpIleLys.....AspL 439
||| ||| ||| ||| ||| |||
2580 CATTTCAAGCTGCCCGCCGACATC.....ATCAAGCGTGAATACGATA 2623

```



1179 taspHisSerGluasnGluaspThrMetSerSerProLeuProGlyL 1196  
1178 TAATGAGGGTGAACAAACAGATTATGCTTGAATGTT 1142  
1196 yLysSerAspLysArgAspSerAspLeuValArgSerGlu 1210  
1141 .....GACCAAGCTCTCCAGTAAGGATGAGTGCCA 1109  
1211 .....LeuGluLysProArgGlyArgLysThrProValThrG 1224  
1108 GGAGACGGCATTCAGATCCCGAGTGGGACTATCTTCTCAATGTTCAA 1059  
1224 uGlnGluLysLeuLysMetAspAspLeuThrLys 1236  
1058 GCAGACAGAAAAATCAAAAGTTGATGCTTGGCTAGTGAAGATGAAC 1009  
1237 .....LeuValGlnGluLysProLysGlySerGlnArgSerArg 1250  
1008 ATGGAACGGTGGCCAGAGAAACATCAAGTGGTGCATTAACAGTAGA 959  
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluG 1267  
958 AACAAAAATAATGACAGGATGGACAGCACACCCATTCAGGCACCTGAAGA 909  
1267 uLysArgLeuLysGluAspLeuLysGluAsn 1278  
908 AGAATCTAGGAGAGTGTATATGACCGAAGAGGTATCAATGTACGCTTAG 859  
1278 luAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPro 1294  
858 AAGATAGTGAAACATCGCAACTTAATAAAATTTGAAGGCA 818  
1295 LysProLeuGlyGlyGly 1300  
817 ...CCGCTTAAGGAGAGATAAGGATGCAATATTAATGACGATTTGTGACC 771  
1301 .....ThrProLysGluGluProThrMetLysThrSerLysLysG 1314  
770 TGTGTAAGAAAATGTCAGAGATGAGGCTTTTAAATAATCAATGGAACATG 721  
1314 LysSerLysLysSerGlyProProAlaProGluGluGluGluGlu 1330  
720 CGGTACTAAGAGACAGGT 686  
1331 Glu 1331  
685 GAAGTTGCTGATATTTCAGGTTCACGAGAAATTCGTATACCCATTTTCAA 636  
1332 .....ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnH 1346  
635 TTCTTTGAAATCAAGGGTCAAGAGCCCAAAATAAAGAGAAAATGAAAA 586  
1346 isArgValSerArgArgAlaGlnGluArgAlaGluSerProGluSerSer 1362  
585 AGAGTTGCAAGAGATGAGTTAATGCCACACAGATTCACCTCTCTCGAATG 536  
1363 AlaileGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSe 1379  
535 ACA...GAAAAATCAAAATATCAACGGCGACAGATGCTTAGACACCGGTACC 489  
1379 LysThr 1381  
488 AAAAACA 482

seq\_name: gb\_p13:SCRIF1

seq\_documentation\_block: 6541 bp DNA 11-AUG-1992  
LOCUS SCRIF1 PLN  
DEFINITION S.cerevisiae RIF1 gene.  
ACCESSION X66501

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
Source

gene  
CDS

X66501.1 GI:4327  
RIF1 gene.  
Saccharomyces cerevisiae.  
Saccharomyces cerevisiae.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
Hardy, C.F., Sussel, L. and Shore, D.  
1 (bases 1 to 6541)  
A RAPI-interacting protein involved in transcriptional silencing  
and telomere length regulation  
Genes Dev. 6 (5), 801-814 (1992)  
2 (bases 1 to 6541)  
Shore, D.M.  
Direct Submission  
Submitted (30-JUL-1992) D.M. Shore, College of Physicians &  
Surgeons of, Columbia Univ., Dept. of Microbiology, 701 West 168th  
Street, New York, N.Y. 10032, USA  
Location/Qualifiers  
1. 6541  
/organism="Saccharomyces cerevisiae"  
/strain="AB320"  
/db\_xref="taxon:4932"  
454..6204  
/gene="RIF1"  
454..6204  
/gene="RIF1"  
/codon\_start=1  
/protein\_id="CAA47121.1"  
/db\_xref="GI:4328"  
/translation="MSKDFSDKKHTIDRQDILRRSHQDNYSNGSPWMTNLP  
SPOAHMHQSDLSPTPKRKLASSDCENKOFDLSAINKNLYPEDTSGRLMSQSL  
ASNDNVSPVTKSFAFSRIEPIYRPGSPKSPGKSLRNKLPVSRVSD  
LSYNKQYVTKLHNGNFTSPYKTRVNPRALEYVSGETHGLDNVESVSFKFIE  
GGLIARQEDSYARREVEATFNIIPIILTKMNVNQDQFNLIIVNIESIIEICI  
PHLOIAOTLLSSSEKPNFVIRLVQIVRFSAIMSNFKIVKTKRPLDNVNLKVI  
RYWTTGALRNSNKIITTAQVSFLRDEKFGTFILSNEIKPILSTFTEIMFNSHL  
IYEKLLIRGFLSKYPLMIETVTSWLPCEVLPRIIGDEIYSMLITISVIVLELL  
KKDFVDEHERIYOCIMLSPVCTIPEKFLSKPLNSDLSANLKVTHIGHLTQOIK  
NIYVKNDKIAMDLWSMTGLLYDSGRVYDLTSESNKVFMDLNNLCFINHPTRL  
MSIKVMRIITYCISTKISKQNGKNSLLSLRTPFTQMTLYPNDSAREGIYHLLG  
VYTAFTSNKLSLDMFELFWDHLITPIYEDYVFKYDSTHQNVLTVHLLIGGKNA  
DVALERKXKHHPMSVIASGVKLDDISLIPQIIRKREYDKIMKVVFAVATSNV  
NLADHLILSLKHLPEKDDOHLFESLSILKVTQNNKDTPIFROFGAVTSSVYT  
FDLFLRKNDSLVNFTQISKVGISQGNWLDLLKDVIRKARNETSEFLIEKLEL  
DKKTEVYAQNWVGSTLUPPNISFREFOSLANIVNKNPNENSIEFIDLCLSPVN  
LFTLLHVSMWNNNFYFIQSYVSKNENKLVNLDLTLKTSLPNGPELFGSLPLVR  
NKFMDELYCIHSPNLLNSIPDLNSDLLKLLPRSRASYFAANIKLFCSQSLTVR  
WLLKGOOLEOLNPFSEIENVLONASDSELEKSEIIRLLHLAMANPEPLFGLLNF  
CIKNMADHDEFCCGNTSEVLFKLSPELLKLLTYKPKNGKLLAAVTEKLENGDD  
YILELEKIIIOKEIOILEKULEPLVFLNPNVSMOKHKKASTNMLRELVLVLYTKP  
LSRAAKKFFSMLISLIPPNPYQTDIMVNLIDLIKSHNRKFKDKRTYNATLKTKG  
WIOESGVVHQGDSKEIEAIPDTKSMYIPCEGSEKNLNLQKRVSDIQVDPATQMG  
EPPSIOISSOISAKNDSISLKNATIMNSOQESHANSRSIDDETLEEVDSNRE  
IDQOMKSTOLDKRVANHSICSTKSDVDVTEHESIDTOSSEVNAVOPLEVLTSELK  
AVTNRSIKTNDPNVNSDNPLKRPSEKPTSENKRSKRGHETHVDVLYSEEOAVSPSS  
DVICTNKSANESSESLALNSIKVETCNKARNPNENGNNDYACNVQSPVREYVPGD  
ROENOSMNKINSKFTQDNTAQYKSKARNPNENGNNDYACNVQSPVREYVPGD  
GTOIPSGTILLNSSKQKVDLRSDEDEHGTVAQKHQVGAISNRKNRNDMDST  
PTQGTESREVMTEEGINVRLEDGTCCLNKLKPLKDGKNDANINDFPVPEVNV  
RDEGLKSMHSAVSKETGLEQPEVADISVLPETIPFNSLKMGGSKSOIKEKLLKR  
LORNELMPDPSPPRTMTNTINAOGLDTPVKTIGCKEKKHHEIQLOAHTADGEP  
GGGDNEDATREATPSLKVHFFSKKSKRLARLGRFTPGDLNGISVEEERNRURIELL  
FMRLEYYSNRNDNMN"

BASE COUNT 2340 a 1116 c 1224 g 1861 t

alignment\_scores:  
Quality: 226.00 Length: 1819  
Ratio: 0.292 Gaps: 80  
Percent Similarity: 42.496 Percent Identity: 17.372



```

736 .....ilePheSerLysGlu..... 741
    ::::||||| ::::|||||
2887 AACATAAAACTATTCAAGTCTCCGAACAGCTTACATTAGTTGGT 2838
742 .....ThrGlnPheAlaGlnIleP 748
    ::::||||| ::::|||||
2837 GTTGAAGGTCACAACTTGAACAGCTAAATCAAAATTTTTCAGAAATCG 2788
748 heGluProLeuHisLysSerLeuAspProSerAsnLeuGlu.....His 762
    ::::||||| ::::|||||
2787 AGAACCTTTTACAAATCGCTGAT...AGCGATTAGAAAATCAGAA 2741
763 LeuIleThrProLeuValThrIleGlyHisIleAlaLeuAlaProAs 779
    ::::||||| ::::|||||
2740 ATAATCAGAACTCCTT.....CATCTGGCATGCAAAATCCCAT 2700
779 pGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysA 796
    ::::||||| ::::|||||
2699 TGAA.....CCATTATTCAGTGGCTATTGGAATTTTTCATAAAGA 2659
796 splLeuLeuMetAsn.....AspArgLeuProGlyLysLysThrLys 810
    ::::||||| ::::|||||
2658 ATAATATGCGGATCATTTGGATGAATTTTGGGAAATATGACCACGAA 2609
811 LeuTrpValProAspGluValSerProGluThrMetValLysIleG 827
    ::::||||| ::::|||||
2608 GTTCTTTTC.....AAAATAAGTCCAGAGCTACTGTTAAATTAAT 2568
827 nAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnHis 844
    ::::||||| ::::|||||
2567 GAGGTATAG.....GAGA 2554
844 erLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
    ::::||||| ::::|||||
2553 AGCCTAATGGAGTTACTGGCAGCAGTATTCAAAAATTTGAATGGA 2504
861 AspGlyAsp.....LeuThrGluGlnGlyLysIleSerLysPr 873
    ::::||||| ::::|||||
2503 GATGATGATTATATATTGGAGTTGCTGGAAGAGATTATTATACAAAAGA 2454
873 oAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
    ::::||||| ::::|||||
2453 G.....ATCCAAATTTTGGAAAAAT 2434
890 laGlnGluProCysTyrHisGluIleThrLeuGluGlnTyrGlnLeu 906
    ::::||||| ::::|||||
2433 TAAAGGAGCA.....CTTTTAGTTTTCCTTCA 2405
907 CysAlaLeuAlaIleAsnAspGluCysTyrGln.....Va 918
    ::::||||| ::::|||||
2404 AATCCAGTTAGCTCTAATATGCAAAAACACAAAAGAGTACAAATATGCT 2355
918 lArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgL 935
    ::::||||| ::::|||||
2354 CCGTGAATTAGCTTATTATATCTACTAAACGCTTTCTAGAAAGTCCG 2305
935 euProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro... 950
    ::::||||| ::::|||||
2304 CCAAAAAATTTTTTCCATGTTAATTAGTATATTACCGCCAAACCCCTAAC 2255
950 ..... 950
2254 TATCAAACTATTGATATGTTAATCTGTTGATTGACCTCATCAAAATCACA 2205
951 .....ValLysGluArgArgAlaHisAlaArgGlnCysLeuValL 964
    ::::||||| ::::|||||
2204 TATCTGTAGTTTAAAGATAAGAACCTTAT.....AATGCAACATTAA 2161
964 ysAsnIleAsnValArgArgGluTyrLeuLysGln.....His 976
    ::::||||| ::::|||||
2160 AGACAATAGGA.....AAGTGGATACAGGAATCGGGCGTGGTTTCAT 2120

```

```

977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGlu..... 989
    ::::||||| ::::|||||
2119 CRAAGGGACTCAAGTAAGGAAATGAAGCTATACCTGACACTAAGTCGAT 2070
990 .TyrVal..... 991
    ::::||||| ::::|||||
2069 GTATATTCCATCGGAAGGAGTGCAAAATAAATTTTCAAACTTACAGAGAA 2020
992 .....ValProTyrThrIleHisLeuLeuAla 1000
    ::::||||| ::::|||||
2019 AAGTGGATTCTCAAGATATCCAGGTTCCGGCTAG...CAGGGAATGAAA 1973
1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysaspVa 1017
    ::::||||| ::::|||||
1972 GAGCCCTCCCTCCTCAATTCAAATATCTCCAGATTAGTCGGAAGAT.. 1925
1017 lLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluA 1034
    ::::||||| ::::|||||
1924 .TCAGATTCAATATCGCTTAAANAATACTGCAATAATGAATTTCTTCACAGC 1876
1034 snAsnSerHisAla..... 1038
    ::::||||| ::::|||||
1875 AAGAATCACATGCTAACCCGAAGCAGATCAATTGATGATGAAACACTAGAG 1826
1039 .....PheIleArgLysMetValGluAsnIleLysGlnTh 1050
    ::::||||| ::::|||||
1825 GAAGTGCATAATCAAAAGCATTTAGAGAAATAGATCAGCAGATGAAGATAC 1776
1050 rLys.....AspAlaGlnGlyP 1056
    ::::||||| ::::|||||
1775 GCAGTTAGATAAAACGTGGCCAAATCATAGCAACATTTGTTCTACTAAAA 1726
1056 roAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
    ::::||||| ::::|||||
1725 GCGATCAGGTGGATGTTACTAG...CTGCATGAAAGTATTGATACACAA 1679
1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLy 1089
    ::::||||| ::::|||||
1678 TCCTCGGAAGTGAACGCATACCACCGCATAGAAGTTCTTCATAGCGAACT 1629
1089 sAspProValLeuProAlaArgPheThrGlnProAspLysAsnPheS 1106
    ::::||||| ::::|||||
1628 GAAGCGGTAAACAAATAGATCTATCAAAACGAATCCTGATCATAACGTTG 1579
1106 exAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrPro 1122
    ::::||||| ::::|||||
1578 TTAACAGTGATAATCCTCTCAAAACGACCTTCCAAAGAGAGCGCTACCTCT 1529
1123 GlyLysProLysThr..... 1127
    ::::||||| ::::|||||
1528 GAAATATAAAGATCGAAAGGTCTATGAACGATGGTGGACGTTTTAGTTTC 1479
1128 .....ThrAsnValL 1131
    ::::||||| ::::|||||
1478 TGAGGAAACAAAGCGGTGTCGCCCTAGCAGTACGCTTATATGTACTAATCA 1429
1131 euGlyAlaValAsnLysProLeuSerSerAla..... 1141
    ::::||||| ::::|||||
1428 AGAGTATAGCCAAACGAAGAATCTTCGTTAGCTTTAAGGAATAGCATAAAA 1379
1142 .....Gl 1142
1378 GTGGAACAAACTGTAAATGAAATTCCTTTGAATGTCACCTTTAGATCTTGA 1329
1142 yLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlas 1159
1328 CCAGCAAAACCAATAACAAAGAGATCGAAAGCGGAAGTTGAACATGTTTC 1279
1159 exSerSerSerAsnProSerProGlyArgIleLysGlyArgLeuAsp 1175
1278 AAAGACAGGAGAACCGAGAAAGTATGTAATAAATCAACAGTAAAGTTTTT 1229
1176 SerSerGlu.....Me 1179

```



3682 .....TTGAAGCATCTT 3671

506 valLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522  
: : : : :  
: : : : :  
3670 CCAAGGAT .....AGAAGGCCAAACTCATTT 3642

522 llyslalalpheSerLysValMetValIleThrArgAsnLeuProAspp 539  
: : : : :  
: : : : :  
3641 AGAATCCCTTTAGCAGTCTATATTAAAGTCCAGCAAAACAACAAGATA 3592

539 roGLyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu..... 553  
: : : : :  
: : : : :  
3591 CACCATTTCGTGACTTTTTTGGTGCTGTACTTCATCGTTGTTTAT 3542

554 .....GluAspAspGluLysIleAlaGLy 561  
: : : : :  
: : : : :  
3541 ACATTTTGGACCTGTTTTTAAAGAAAGATCATTCATATTAGTAAATTT 3492

561 sGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluG 578  
: : : : :  
: : : : :  
3491 CAATATCCAAATTTCCAAAGTTGGTATTTCGCAAGGAACATGACCCCTG 3442

578 lYcysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThr 594  
: : : : :  
: : : : :  
3441 ATCTTTTAAAGATGTTATCAGGAAGCTCGAAAC.....GAAACA 3401

595 AsnProPheLeuGluMetIleLysPheLeuLeu..... 605  
: : : : :  
: : : : :  
3400 TCTGAATTTCTGATATAGAAAGTTTCTAGAAATGGATGATAGAAANAC 3351

606 .....GluArgIleAlaProValHisI 613  
: : : : :  
: : : : :  
3350 AGAAGTATATGCCAAANATGGTGGATCAACGTATTACCACCAATA 3301

613 leAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIle 629  
: : : : :  
: : : : :  
3300 TATCTTTCAGAGAAATTTCCAGATTTGGCGAATATGTGCAACA..... 3257

630 AspGLYThrAlaAspGluAspGluGlyValProThrAspGlnAlaI 646  
: : : : :  
: : : : :  
3256 .....GTCCGAATGAAATTCAT 3237

646 eArgAlaGlyLeuGluLeuLysValLeuSerPheThrHisProIles 663  
: : : : :  
: : : : :  
3236 TCAAAATTTCTGGACCTATGTTTGAAGCTCAGCTT.....CCAGTTA 3193

663 erPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMet 679  
: : : : :  
: : : : :  
3192 ATTTG.....TTTACACTACTTCATGTCGTCTGTGTGTCG 3158

680 AspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGI 696  
: : : : :  
: : : : :  
3157 ACAATAATTTTATTATTTCATTCAAAGCTATGTTTCAAAAAATGAA..... 3110

696 ySerLysIleGluGluAspPhe..... 703  
: : : : :  
: : : : :  
3109 .AACAAATTAACGTGGATCTGATCAGCTGTTAAARACTTTCACAG 3061

704 ....ProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSer 718  
: : : : :  
: : : : :  
3060 GGAATCCTGAACGTGTTTCCGCTGCTGCCATTTTAA..... 3023

719 LysLysGlyProProArgGlnAlaLysIyr.....AlaIleHisCy 732  
: : : : :  
: : : : :  
3022 .....AGCGTAACAAATTTATGATATTFTTGGAGTACTG 2988

732 sIleHisAla..... 735  
: : : : :  
: : : : :  
2987 TATTATTCCAATCCAAATTTTACTTAATTTCCATCCCTGATCAATTCAG 2938

735 ..... 735

2937 ATCTCTATTGAAGCTGTTTACCACGTAGCAGAGCTTCTATATTTTGCAGCA 2888



MEDLINE 95112788

## FEATURES

```

Location/Qualifiers
1. .6449
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="II"
/complement(182..5932)
/gene="RIF1"
/complement(182..5932)
/gene="RIF1"
/feature="RIF1"
/feature="ORF YBR275c"
/codon_start=1
/protein_id="CAA85238.1"
/db_xref="GI:536715"
/translation="MSKDFSDKKKHTDRIDOHILRRSDHNSNGSSPMWKTNLPP
SPOAHMIQSDLSPTPKRRKLASSDCENKOFDLSAINKNLYPEDTGRSLWQSPPELS
ASNSDNPVPTKVSFSDRIESSPIYRPGSSPKPSPSSKPKGKILRNRLPSVTVSD
LSYNLQTHKLRNGNIFTSPYRETRVNPRALETWVSGEIHGLVDNESVSEKEIECI
GGLGILROESDYVARFEVYATFNIIPIITTKNVNEVDQKFNILINVTESIEIECI
PHIAQDPTLLSSSEKNPFVIRLYQIVRFSAIMSNFKIKWLTFRPDLVNLKLV
IYKLLIRGFLSKYKPLMIETVTSWLPGEVLPRIIIGDEIYSMKILITSIVLLELL
KKCLDFVDEHRIYQCIIMLSVPCETIPEKFLSKPLASYSANLKYTIHLLTQOIK
NYIVKNDKIAMDLWLSMTGLLYDSGRVYDLSSEKNWFDLNNLCFINHPKRL
MSIKVRLITICIKISQKNOEKNLSLLRTPFQMTLPVNDPSAREGIIYHLG
VYVFTSNKNSLDMFEWDLITPIYEDYVFKYDSIHLQNLFTVLHLLIGKNA
DVALERKYKHHPMSVITASBGVKLKDISSLPPOI IKREYDKIMKVVFQVTEVAISV
NLADHLIITSLKHLPEKDKQTHLESFSSLLIKVTQNNKPTPIFRDFGAVTSFVLT
FLDLFLRKNDSLVNFIOISKVISQGNWTLDLLKQVIRKARNETSEFLIIEKLEL
DKRTEVTAQNWGNTLFPNISRFEOSLANIVKVPNENSINENFLDLCKLSPFVN
LFTLLHVSMMNSNFIYIOSYKSNENKLNVDLITLLKLSLPGNELFGLLPLFR
NKFMIDILCYTHSPNLSNIPDLNLLKLLPRSRASYFAANIKLFKSEQTLVR
WLLGQOLEQNLQNFSEINVLQNASDSELEKSEIIRELLHMANPIEPLFSGLLNF
CIKNMADHLDQFQNTSEVLFKISPELLIKLITYKEKPNGLKLAIVIEKIENGDDO
YIELLEKILIQEIQILEKLEPLLVFNPSNMOKHKKSTMLEULVLLYLRP
LSRSAAKKFSMLISILPPNPYQOTIDMKNLLDILKSHNRKFKDKTYNATLTKR
WIPPSQVHQSQSKLEAIPDKSMYIPCEGSENKLSNLRQKVDSDQIQVATOGMK
EPSSQIISQISAKSDSISLKNATIMNSQOSKHANRSIDDETLEEVDNBSIRE
IDQOKSTQLDKNVANSNICTSKDEVDVTELHESIDTOSSEVNAVQPIEVLTSK
AVTNRSKTNPNDHNVNSDNLKRPSEKPTPTSENKSGKHETWVDVILVSEQAVSPSS
DVQNTKSTANESSILALRNSIKVETNCNENSLNVLTLDDOQITIKEDKGVEHVQ
ROENQSNKINSKFTQDNIAOYKSVKKARPNNENNDYACNVEQASVNRNVPGD
GIQIPSTILLNSKQTEKSKYDLDLRSDEDEHGTVAQEKHQVGAINSRKNRMDST
PIQTEESREVMTEGINVRLEDGTCNLNKLGLKGDKNANDINDPVPVEENV
LQNEGLKSMHAEVSKETGLEQPEVADISVLPEIRIPIFNSLKMQSGSKQIEKILKR
LQNELPPPOPPMTENTININAQGLDTPVKTIGGKKEHIEIQLOQAHTBADGEPLL
GGCNGEDATSEATPSPSLKVHFFSKSRRLVARLGRFTPGDLNGISVEERNLRITELD
FMRLSEYYSNRNDMN"
misc_feature
3859..3869
/feature="ARS-consensus"
BASE COUNT 1829 a 1213 c 1086 g 2321 t
ORIGIN

alignment_scores:
Quality: 227.00 Length: 1819
Ratio: 0.293 Gaps: 80
Percent Similarity: 42.551 Percent Identity: 17.372

alignment_block:
US-09-512-581-2 x SCYBR275C/rev ..

Align seg 1/1 to reverse of: SCYBR275C from: 1 to: 6449
3 HisSerLysThrArgThrAsnAspGlyLysIleThrTyr..... 15
5386 TATAAGCAACACAGGTAATCTCCGCCCTTAGACTATGGTGAGTGG 5337
16 .....ProProGlyValLysGluIleSerAspL 25
5336 TGAGATCCATGGTTTAGTGATAATCAAGAGTGTTTCGGAGTTTAAAGAAA 5287

```

```

25 ysile.....SerLysGluGluMet 31
   |||
5286 TTATTGAGGCGGATTGGGCATCTTCGCACAGAATCTGAAGATTACGTC 5237

32 ValArgArgLeuLysMetVal.....Vally 40
   |||
5236 GCGAGGAGGTTTGAGGTATACGCAACCTTTAATAATTATATACCCATTCT 5187

40 sThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLysTyrL 57
   |||
5186 AACCACTAAATAATGTTAACGAGGTTCATCAGAAATTCATATCTTAATTC 5137

57 euAsnLeuAla.....LeuHisLeuAla 64
   |||
5136 TGAATATAGAAAGTATCATTTGAAATATGCATACCCGATCTACAAATAGCA 5087

65 SerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAl 81
   |||
5086 CAGGACACATTTATTAAGTAGTAGTGAAAGAAAAAATCCATTTCGTAATC 5039

81 acysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrT 98
   |||
5038 .....AGACTATATGTTCAATAGTTCGATTCT 5011

98 hrSer..... 99
   |||
5010 TCAGTGCTATTATGCTTAATTTCAAAATCGTTAAGTGGCTAACGAAAGG 4961

100 ProAsp.....LysLeuLysAspIlePhe..... 107
   |||
4960 CCAGATTGGTTAATAAGTTAAAGTTATTACAGGTGGACACCGGTGC 4911

108 .....MetPheIleThrArgGlnLeuLysG 116
   |||
4910 GCTAAGAAATGAAATTCACCAAGATAATCATACACGACACAGTCTCTT 4861

116 LysLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
   |||
4860 TCTTGAGAGACGAAAAA.....TTTGGTACTTTTCTTGGAGCAAC 4820

133 GluAsnIle...AlaTrpValLysSerTyrAspIleCysPheGluLeuG 148
   |||
4819 GAAGAAATTAACCTATCATAGGACCTTCACCTGCAATATGGAATCAA 4770

148 uAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValI 165
   |||
4769 TAGTCAACACCTAATATGAAAAATTG.....CTAC 4738

165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181
   |||
4737 TTATAAGGGGATTTTGGACCAAGTATCCAAAAATTAATGATTGAGACGGTA 4688

182 SerSer.....IleIleCysGluGlyAs 189
   |||
4687 ACTTCTTGTTACCAGTGAGGTACTACCTAGGATTATTATTCGGAGATGA 4638

189 pThrValSerGlnGluLeuAspThrValLeuValAsnLeuValProA 206
   |||
4637 AATTTACTCCAGAAATCTCATCACTTCAATAGTTGTTTACTG.... 4592

206 laHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeu 222
   |||
4591 .....GAACCTACTTA 4583

223 LysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnG 239
   |||
4582 AAAAAATGCTTAGATTTTGTGATGACATGAAGAGATTAC.....CA 4539

239 nValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValP 256
   |||
4538 GTGATCATGCTATCTCCAGTATGCGAAACAAATCCCGGAAAAATTTTAT 4489

256 heAspLeuIleLeuGluLeuTyr.....AsnIleAsp..... 266

```



992 lProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValG 1009  
 89153 .....  
 1009 lnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeu 1025  
 89166 AGGATATTGAAATAATAAAAAAT .....  
 1026 GluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLy 1042  
 89189 .....  
 1042 smetValGluAsnIleLys...GlnThrLysAspAlaGlnGlyPro... 1056  
 89227 ATTAAATAGATCATATAAAATACATATGAAATGAAATACAGCAAAATAT 89276  
 1057 .....  
 89277 TAATAATTGATGATGCCAAAAGAAAGTAAAG .....GAA 89311  
 1071 ValAlaMetAsnIleIleMetSer...LysSerThrThrTyrSerLeuG 1086  
 89312 ATAACAGATAATATTAACAAGCGCTTTTAAATGAAATACAGAAAATATATA 89361  
 1086 userProLysAspProValLeuProAlaArgPhePheThrGlnProAspL 1103  
 89362 TAATGAAATAATGCGGTAAATAATCTGCA .....A 89393  
 1103 ysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhe 1119  
 89394 AAAATATTGCGTAAAGCTACTTATTAAATAATGAATTAGATAAATTT 89443  
 1120 PheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValasnLy 1136  
 89444 TTATTG .....AAATGAATGAATTTAATGCTATAATAATA 89481  
 1136 sProLeuSerSerAlaGly .....LysGlnSerGlnT 1147  
 89482 TGATATAAGATCTGCTGTGATAAATAATTAATTAAGAAGAAGAAG 89531  
 1147 hrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsn 1163  
 89532 AAAGAAAGAAAGA...GAAAGATTGGAAGAAAGCGAAACAAAGAAGAA 89578  
 1164 ProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetas 1180  
 89579 AGAAAAGAGAGAGAAGAATAAGAAAAAGAAAAACAAGAAGAAGAGACT 89628  
 1180 pHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysL 1197  
 89629 GGAAG 89678  
 1197 ysSerAspLysArgAsp .....AspSerAspLeuValArgSer 1209  
 89679 AGCAAGAAAGACAAAGCAACAAACAAAGAAAGAGAGAGAGAGAGAG 89728  
 1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGluGlnG 1226  
 89729 GAACAAGAA .....CGACTACAAAAGAAAGAAATTAATAAGACAGA 89772  
 1226 u...GluLysLeuGlyMetAspLeuThrLysLeuValGlnGlnL 1242  
 89773 CGAAGAAAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 89822  
 1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258  
 89823 AATTAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 89872  
 1259 SerAspGluGlnGlnTrpProGlu .....GluLysArgLeuLysG 1272  
 89873 TTAGAAGAGCAAAAAAGCGCTGAAATAATAATAATGAAGCAATGGTAAAGGG 89922

1272 uasp...IleLeuGluAsnGluAspGluGlnAsnSerProProLysLysG 1288  
 89923 GGATAAATACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 89971  
 1288 lyLysArgGlyArgProProLysProLeuGlyGlyThrProLysGlu 1304  
 89971 ..... 89971  
 1305 GluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyPr 1321  
 89972 ...CCTAACGTTAGTAGGATACTAATAATAATAGTCCAATAGTAACAG 90018  
 1321 oProAlaProGluGluGluGluGluGluGluGluGluGluGluGluG 1338  
 90019 TGAATTACAGAAAGCGATGATATTGATAACAGTCGAAATATACATACTA 90068  
 1338 luGln .....LysSerLysSerLysGlnHisArgValSerArg 1350  
 90069 GTCATATGAGTGACATCGAAAGTACACAAACTAGTCATAGAAGTAACACC 90118  
 1351 ArgAlaGlnGlnArgAlaGluSerProGlu 1360  
 90119 CATGGCAACAAATCAGTGATATTGTTCAA 90148  
 seq\_name: gb\_pl4:SCYBR275C  
 seq\_documentation\_block: 6449 bp DNA PLN 11-AUG-1997  
 LOCUS SCYBR275C  
 DEFINITION S.cerevisiae chromosome II reading frame ORF YBR275C.  
 ACCESSION Z36144 Y13134  
 VERSION Z36144.1 GI:536714  
 KEYWORDS  
 SOURCE Saccharomyces cerevisiae.  
 ORGANISM  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 REFERENCE  
 1 (bases 1 to 2920)  
 AUTHORS Aigle,M., Baclet,M.C., Barthe,C., Biteau,N., Crouzet,M. and  
 Daignon,F.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 2644 to 6449)  
 AUTHORS Brandt,T., Christiansen,C., Holmstroem,K. and Kallese,T.  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 6449)  
 AUTHORS MIPS.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-AUG-1994) Data collected by MIPS on behalf of the  
 European yeast chromosome II sequencing project. MIPS at the  
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152  
 Martinsried, FRG; E-mail: Mewes@mips.embnat.org  
 REFERENCE 4 (bases 1 to 6449)  
 AUTHORS Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,  
 Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T.,  
 Brendel,M., Brueckner,M., Bussefau,F., Christiansen,C.,  
 Contreras,R., Crouzet,M., Czepluch,C., Demolis,N., Delaveau,T.,  
 Daignon,F., Domdey,H., Duesterhus,S., Dubois,E., Dujon,B., El  
 Bakoury,M., Entian,K.D., Feuermann,M., Fiers,W., Fobo,G.M.,  
 Fritz,C., Gassenhuber,H., Glandsdorff,N., Goffeau,A., Grivell,L.A.,  
 de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,  
 Jacq,C., Jaquet,M., Jauniaux,J.C., Jonniaux,J.L., Kallese,T.,  
 Kieau,P., Kirchraht,L., Koetter,P., Korol,S., Liebl,S., Logghe,M.,  
 Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L.,  
 Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,  
 Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,  
 Pohl,F.M., Pohl,T.M., Potier,S., Profit,M., Purnelle,B., Ramezani  
 Rad,M., Rieger,M., Rose,M., Schaeff-Gerstenschlaeger,I.,  
 Scherens,B., Schwarzlose,C., Skala,J., Slonimski,P.,  
 Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R.,  
 Urrestarazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A.,  
 Vetter,I., Vierdeels,F., Vissers,S., Wagner,G., de Wergifosse,P.,  
 Wolke,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Klein,K.  
 Complete DNA sequence of yeast chromosome II  
 EMBO J. 13 (24), 5795-5809 (1994)  
 TITLE  
 JOURNAL







302 AlAlYsAspSerGluLeuAlaSerGlnAsnLysProLeuTrp..... 315  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
86702 AAAAAAAAAAATTTCGAATCTAATAACAACATCATTAACACTACTTAAT 86751

316 .....GlnCysTyrLeuGlyArgPheAsnAspIleHisV 327  
: : : : : ||| :  
86752 GGATTCATTCAGACTGATGTTTATTAAGTAATATAAATGATTATAATA 86801

327 alProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsn 343  
: : : : : ||| :  
86802 TAAATGAATAATTGAA.....AAA 86821

344 HisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerH1 360  
||| :  
86822 CATCAAAATATATTGAATGAAATATATTAATGGATTTAATGAA..... 86863

360 sAspProGluGluAlaIleArgHisAspValIleValSerIleValThrA 377  
:  
86864 .....TCATATAATATTATTAATAACAATAATGACTG 86894

377 laAlaLysLysAspIleLeuLeuValAsnAspHisLeu..LeuAsnPhe 392  
: : : : : ||| :  
86895 AA.....ATTATAATGATAATTTAGATTATAATGAA 86926

393 ValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMe 409  
: : : : : ||| :  
86927 ATAAAGAA.....ATTAAAGAAGTAGCAC 86952

409 tMetClyLeuAlaGlnIleTyrLysLysTyr.....AlaL 421  
: : : : : ||| :  
86953 AACAGAATATGATAACCTAATAAAAAAGTTGATGAATAAAAAATTATT 87002

421 euGlnSerAlaAlaClyLysAspAlaAlaLysGlnIleAlaTrpIleLys 437  
:  
87003 TGAATAATTTAAAGAACAGAGAGACATCGATTAAITGATTATATAAA 87052

438 AspLysLeuLeuHisIleTyr.....TyrGlnAsnSerI1 449  
:  
87053 GAAAAAATATTAACTTATATAATAAAATGTTCAGAACACAAATATAAT 87102

449 eAspAspArgIleuLeuValGluArgIlePheAlaGlnTyrMetValProH 466  
||| :  
87103 AGATGATTCTTATATTAATATACAGTTAAAAAACAGTAT..... 87142

466 lAsnLeuGluThrThrGluArgMetLysCysLysTyrTyrLeuTyrAla 482  
: : : : : ||| :  
87143 .....ATTAAACATTTGGAGATGGAAATTTTA..... 87172

483 ThrLeuAsp...LeuAsnAlaVal.....LysAlaLeuAsnG1 494  
||| :  
87173 ...TTAGATTCTTCAACACATAGACAAAAAATAAATCAGTAGCAA 87219

494 uMetTrpLysCys.....GlnAsnLeuLeuArgHisG 505  
: : : : : ||| :  
87220 TCTAGAAATTTGCTACTAATAAAGAGATATAAAAAATTTACTTAAACATG 87269

505 lNValLys..... 507  
:  
87270 TTATAAGTTGGCAAAATTTTCAGGTATTATGTAATGCTGTGATACAAAT 87319

508 .....AspLeuLeuAspLe 512  
||| :  
87320 ACGGAAATAACTCCAGAAAAATCCITTAGAAGTAATGATTATTAATTT 87369

512 uIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysV 529  
| :  
87370 A.....CAATTATATTTGAAAGAA 87389

529 alMetValIleThrArgAsnLeuProAspProGlyLysAlaGln..... 543  
||| :  
87390 AACATGAAATACACATTCACATTTGGAAAAATGATCTCGATTAGAGTTAGAT 87439

544 .....AspPheMetLysLysPheTh 550







```

723 .....ProArgGlnAlaLysTyrAlaIleHisCysIle 733
8537 TGAGATCCAAGCGTTTCACAGACACACGGAA...GTTCTCCAC..... 8497
734 HisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPr 750
8496 .....AAGGAGCCAAAGACGGTCGAGAGTTCGAACAC 8465
750 oLeuHisLys.....SerLeuAspProSerAsn.....LeuG 761
8464 TATTCGGGAAGAATGATTCATTAAGGCGCAAAATCACAGGATGAATPAC 8415
761 luHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuAla 777
8414 AACAGCCAATAGATCGCAAAATTCT..... 8389
778 ProAspGlnPheAlaLysProTyrLysSerTrpValAlaThr.....Ph 792
8388 ...GACATGTTCTTGAAGCCTTTCGAGAGCTTTCGAGATCGAAGAGGG 8342
792 eileValLysAspLeuLeuMetAsnAspArgLeuProGlyLysThrT 809
8341 AATTCGTTTCAGAAAGAGATATGAAGAAATGTGAAGCAAAATACTA 8292
809 hrLysLeu.....TrpVal..... 813
8291 CCAACTTCAGCTATGCCATGGAACAAATGCTTCGAAACTTGTGAC 8242
813 ..... 813
8241 TATAATGTTAAACATCAAGCGGTACGATTGAGCGATTTTAAGAAGATTAG 8192
814 ProAspGluValSerProGluThrMetValLysIleGlnAlaIleL 830
8191 ACCAAGACGACGAGAGATATGGAACCC.....AAGAGAAATCTATTG 8148
830 yMetMetValArgTrpLeuLeuGlyMetLys..... 840
8147 AGATTGGTGGAAACCATCCAAACACCTCAAGTCGTGTTTGAACGTGG 8098
841 .....AsnAsnHisSerLysSerGly..... 847
8097 CAAGGACATGTTCTATGAAATTTAACTCTCAATCTCCTGATGCAAAAC 8048
848 .....ThrSerThrLeuArgLeuLeuThrThrIleLeuH 859
8047 TTCTAGTGGTGGATCCAAACAAACACTGAAGCTTGTCAAGACAAAGTGCAC 7998
859 isSer..... 860
7997 ATTCGTGGCTGGAACTCACTCCAAAAACTCTTTTACTGTTCCGTCG 7948
861 .....AspGlyAsp.....LeuThrG1 866
7947 CTCTAGAAATTGGACGATTGTCGCCCTCAAGAAATACATGGCTGACGAA 7898
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaG 883
7897 CACAAGAAAAAATTCGAAGGGTGCAGAA.....AAAGTTAAGGAAA 7857
883 lySerAlaIleValLysLeuAlaGlnGluProCystYrHis...GluI1 898
7856 ACATGGACCTGTTTCACAAAGCTCCCGTCCCGTCTCCCAAGAGAG 7807
898 eileThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluC 915
7806 GAAACTCTGAACCAACACCCT..... 7785
915 yTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931
7784 .....CGCGGCTGTTGCTCTTCGACCATTCGCGCTGAAT 7746
932 Arg.....Le 933

```

```

7745 CGAAGAAAAAGTACCGTCGAGAAAGTTGACACTCTCTTCACTCTGACT 7696
933 uArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 950
7695 CCACCTGCCGCTTCGGTCTCTG..... 7674
950 roValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
7673 .....CTCCACCACCAAAATGCGCAGCGATCAACAGT 7641
967 AsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLe 983
7640 GGAGGCGAGCGATGATGCTCGAA..... 7617
983 uLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuA 1000
7616 .....GATATGGGTGCCAGCAAGCGTGTGTTGTA 7583
1000 laHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
7582 GCCACCATCGGCTCTTTCGAAAACTGAATCATTTGAAGGAGCTTCAACA 7533
1017 ValLysGluCysLeu.....TrpPheVa 1024
7532 TTCAGAGTCTGCTTCGACGACTAGGTACACCAGCGTTGGTGGTTTCA 7483
1024 lLeu.....GluIleLeuMetAlaLysA 1032
7482 CATTCCTACTCCAGAACCTGTACAACTAACAAAGATGGCGCAGCTAAAA 7433
1032 snGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLys 1048
7432 ATCCGAACACAGAAAGAT.....ATAAACACACAGTGGCAAAACCAAGG 7389
1049 Gln.....ThrLysAspAlaGlnGly...ProAsp 1057
7388 CAACAGCACCCCTCCGATCACTCCAAAAGCTGCATCAATCAGCGCTGCT 7339
1058 AspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAs 1074
7338 ACTCAAAAGCTGCAGCAATCCGAATTCACAAAGAGTGAAGCTTCCAGAA 7289
1074 nIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAsp 1091
7288 TGTTCATCTCT.....TCTGGAAGTCTCAGAAACAGG 7254
1091 roValLeuProAlaArgPheThrGlnProAspLysAsnPheSerAsn 1107
7253 CAGACGTCCAATCTCGG..... 7237
1108 ThrLysAsnTyrLeuProGluMetLysSerPheThrProGlyLy 1124
7236 .....CCTAAATTACAAAGTGTA..... 7219
1124 sProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerA 1141
7218 ....CAGGAAAAACGCTCTGTGCCAAAAGTGTCAAAACCGTTAGATCCA 7173
1141 laGlyLysGlnSerGlnThrLysSerArgMetGluThrVal..... 1155
7172 AAGTCTCTCAAAAGGACCAAGCAAGCAAAAGAAATGAGAAAGTGGTGCCT 7123
1156 .....SerAsnAlaSerSerSerSerAsnProse 1165
7122 AGTACCAACGCTACTTTGGATGATCAAGCTTCATCTTCTCGAATCAGCA 7073
1165 rSerProGlyArgIleLys..... 1171
7072 AGGACCAAGGATCCGCAAAACCAACATCTGCATCAATTTTCACTGCATTAT 7033
1172 ..GlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyr 1187

```











REFERENCE  
AUTHORS

1 (bases 1 to 40233)  
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,  
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,  
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roodra, A.,  
Saunders, D., Shownkeen, R., Smalton, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
Wilkinson-Sproat, J. and Wohldman, P.

## TITLE

2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*

JOURNAL  
MEDLINE  
REFERENCE

Nature 368 (6466), 32-38 (1994)  
94150718

## AUTHORS

2 (bases 1 to 40233)

## TITLE

Chisoe, S. and Sansone, J.

JOURNAL  
REFERENCE

The sequence of *C. elegans* cosmid F39C12

## AUTHORS

3 (bases 1 to 40233)

## TITLE

Waterston, R.

## JOURNAL

Submitted

## COMMENT

Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1HQ, England  
e-mail: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequenced overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone

## NEIGHBORING COSMID INFORMATION:

The 5' cosmid is R08E3, 2000 bp overlap; 3' cosmid is T07F12, 200 bp  
overlap. Actual start of this cosmid is at base position 1997 of  
CELTF39C12; actual end is at 5117 of CELT07F12

## NOTES:

Coding sequences below are predicted from computer analysis, using  
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES  
source

```
1. 40233
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/complement(212..18270)
/gene="F39C12.1"
complement(join(212..469,519..665,889..1137,1192..1327,
1377..1546,1675..1923,1983..2286,2331..2431,2478..2921,
2964..3128,3178..40758,11107..11265,11315..11701,
11775..12658,12722..13094,13141..13596,13653..13961,
14015..14191,14239..15440,15491..15577,15622..15793,
15845..16066,16113..16227,16276..16725,16780..16875,
17643..17759,17836..18002,18059..18143,18215..18270))
```

## gene

## CDS

```
/gene="F39C12.1"
/note="contains a short region of similarity to troponin T
cardiac isoform; coded for by C. elegans cDNA yk38l1.5;
coded for by C. elegans cDNA yk31c2.5; coded for by C.
elegans cDNA yk44a1.3; coded for by C. elegans cDNA
yk44a1.5"
/codon_start=1
/codon_id="AAB94194.1"
/db_xref="GI:2736399"
/translation="MTDPDGGDIFKIFKICPNEFNEFGLKREMFISERIIVEYDPTD
RLTSKGNLCSLGETIYGLVRDLKGLMAVNRHSTSSVNTVTVDLVPKPAIRNIAH
RKSREIMHPMDAPSESEPVFKHLTKDSKOEFTVTAIDFHEHITVYFVSKNTGICIL
HLSKGNLCSLGETIYGLVRDLKGLMAVNRHSTSSVNTVTVDLVPKPAIRNIAH
YDSVSPNTDIIWHGFFLLAVGDESRLVMDVRCRPAFNAVFDGGVQKATLCSSTRP
LETACWNSNLDYDVRPEVPCAIYDGLKESNIDLMKFNPIINELFVHTNSHVR
FAMNMPHTLIGGRPISEKPYKQVITFKHIDRGCTATRHMLRREKMPYEDVPL
LGGYGOYAAKNDPQVPIPTLWTYPAEVDKFLAKIQDVIDEYSLYKAEVRSVAG
RDKFKHKPLIYQAKGVSEPTDFVPGVNGIMCFORNTFTYKIVRSKPHGVSNPYAG
SDSVFFYRGLQIOEMIQAOTLTMMCNRLAGWNAVQVSESLKELAKPKITESRSWA
KWLNSVIOYGGKEFGNIGVNDLEWEDVPLPGRPLPEVEIPPEYKMDNVAIEEIT
QWNAFFRLFLPDRDLSSESETEHDNAPGNCEDDPEPSGLPSLSLSENMRT
LKCHAKIRYNNKQSDRYMAGNSFAIGSKRIRDKFFKLGKNSITKPKQTLGHR
LKTDAFKKSLDLDLPKFORFKRFPKLPKLTJTDSTSVFMKSEYIKERBAKER
EMREKAAPIQAPIGLSMPKSRKEPKOTPVTDGNEEDPEHARFIRAAIOADNN
SSDDEIVSPOATPTPPATPTPHVQALSVRNEKDDFEVETEDAVEGTKKRMVY
YQATOPGLGRYKPDRLRESTFIRLEHLMYGRRRHERDILYGRFEENDKQSSV
KERTDLKFLSNHMKVGMOTFYDKNFFFSNCKSYNOILRKAIONMRCLGFSOSHET
YFKSVRHRMLCGVNRVMDKEAIEARYLRKDISIETIKLYNHMTFYQAOLP
YKLDHTIFMENFEGDPTFSQAECPIEKKQATKYDENDKFMFLGRLFCEFTSPN
FMHEIENRVPKPKNGLYLEDFEHLAQSPKPLRRNSCPLGLAIESTENSRV
CNEFFHEIYRIIAYRDMCHOREHGVNPHVGYSEVLPBLTAGVLOQFNTHSVNGDYK
VDDILNTFKLEWDSMISDTRDTRGKFRMRVVPKFPQRLQEGIDLPDIYII
EVMRAAKNLHYSIIFNDILRWPGIQLDEGRDMFTCPKFPVRPAEQULSEIDIF
KSPLCPTKAVEFYSEDFTEVLLIGRDSKKNRIPRYNEHINIPEGLIEILKPAFY
GFFKYDRESYRNEEQLEMKNRRELQAIKRIITFRQPVIEEIOELATEENLWE
IDFHAGDDSYNCENPHTRENIKLSVEEEVEDEPKLGSVLTNLFEQDLVKORLGD
LFFRDTSESEGEDEVKNDSPNDISDFWKIFKVTEDAGVNGRDSDELQVORIG
CPTGSPKSSIKIDYNGMVVPLTELTKQTKVTWGEVTEHSVPRIRHNRHREVE
DAKPYRDLISVQLTRVRKSNDFOKTARKNSDPTAVHHIGITPRASLDQFQSDLP
ATCSACIKEPMTKEICIVCHDPTTIFRCTNIVNLHFEPLICGNNIYKFSQNTYEHNN
QYVPIIISIFDQPKLPVDEEFKKLAVNETVIYENTPAVNAVYRAGSAVNTDMW
AETDRCFKTLQDGLAGFEHREPNVMTWRMDEKSERVSEEDVAFDRADADIMW
RDMNTISQDQMDMTYDQTCMRIRRWISYAAATRVOKNEKHYKAAAREEA
RILSGELSVPSLLNELKKAKDEVAATATENDDAKKEERKAYLKKKLPIPEE
TEDPKKMPFAPPKSEWLESSEGEDELPIKSLIKSEKMEKFDITRVHMEQPKS
FEGIATVYKGLFEKRPILKNDIVLVTTRKNRRKPPFPRIQRIPLDESRLNTATVSE
LPVQVNRPDMPKMTVPYNSQRLATTAQETKERALRPPTPTLKLKLYTMKDYV
KAFYHRRKVKSGITLKAARQLRKLKLYGHSKPSHTKAWTHVRQIGLKKSVY
QPKLKEAHLPLGLAKOKNKYIFTYFNPKLEKVLQSRKPSRTSSASDSCYCPVIF
PKVQKRVLEKEDHPTNSELGIIERKKELKREDPPAVLMSSTRSTLKPETGTSVK
REISVGAPALPKADVQTLVSLKIDSQTEPKTESIKQRRHPIFHELEKSSINNS
QGIPIASVPEPSCQLPKRSQKQKHKITTNSSNSNIDSDGMMKDRSSSKQPSVAVRS
KSTMIPHSRLRKQMPACRSFVDSKSSYRKALKIARKSIOKESKVKTKPDDSK
PEVIPSDQPKSQLETSVALPSSAKVSTKSEKDLSPKSTELQELERKFLKNVHNR
RSSRPGIHHORPLHPLDTHETDRKDSRAASQFRFOHMRKRETFRORVLLHKEPR
RSSRLNTIRKNDLSLGNHMKYEQPIDAKISDMFLKPMKLRBSKKYKKN
VKANTTKLQPMKQMLKADYNVIRKRLSDFKIRPKQEDMETKESIEIVV
KPSNLSKLRGGRHSMKFNPSQSPDGKTSAGSNKTLKLVTKYSAHSAGLSTSPKTS
FTNVLLELDDSDSLKTLWLTNTNRKNSGKKVKENTAPVQAPVPVSPREEPTPT
PSPVRLRLRLSASEEKVPVEKVDTPSLTPPAASVAPPKPPKAPQAOQWRAAMRAA
YMGASKPVFVKPSSALSKTESKASNIQESASTRVQPALVSHSTPEVPLTRMAA
AKKSEQEKYTPVAKPKATAPPITPKAASIRPATOKLHESELTKSEASRVSSSGK
SOKADVQSRPKLQSVQEKSVPVKVPPLDQPKQEKESKKNEKVPSTNVLDDQA
SSSNQQPGSAAKPSASITAFLEGRKKRKKMTQSTAEISSIPIPSPENQATSETNR
SSPILPRIKKPTGQLIKRSTSDVPKPVVSGAIVKFDGKPIISNPSYKPMARRR
PKKADSAADSGDLSDEGSHGELKDEENSKQKAGTKVRFQISYEPPEPAPARR
VEIKVPEPTPEISFESLAQOAKOTLRKEASKILEHVPTTEQTSPTASTVPTKSG
SKISFGLSGKGVTPKNTAQTSQTLVLESTSSPELLTPSNPETSNTASTSSKQPK
SLSEKATATQMTQTRGPIKLSHRRQLKIGDPSPLQSGDSSLPRAPQSPKPPQ
VLPEQGVQTYNPPSQRYIYRLGRAWPPPKAKTPLSIRVPQSGVSTSTPAIDLSAKE
VLAPIETKTKMRKPKPMETTLVLDGTTEHFKKLTAEKTLQSLRQKSKRNNKKRT
KTSWIVGCDNDKSTKCMQHNKLEEFRLDKLSITRLQLEKMTAISAVNPRFTDSS
STHKLNRDNPKEFTTKLKLRLANNRQKRSIHSQPKRDVAISAIROVKVRF
QSTGKKMMNPILISGFKPSILKPKPATWLFRLDSEGNKLYMKLGLPSDQKNDKE
```



\* 24376 27231: contig of 2856 bp in length  
\* 27232 27331: gap of 100 bp  
\* 27332 31148: contig of 3817 bp in length  
\* 31149 31248: gap of 100 bp  
\* 31249 33450: contig of 2202 bp in length  
\* 33451 33550: gap of 100 bp  
\* 33551 36754: contig of 3204 bp in length  
\* 36755 36854: gap of 100 bp  
\* 36855 39988: contig of 3134 bp in length  
\* 39989 40088: gap of 100 bp  
\* 40089 42367: contig of 2279 bp in length  
\* 42368 42467: gap of 100 bp  
\* 42468 45775: contig of 3308 bp in length  
\* 45776 45875: gap of 100 bp  
\* 45876 49987: contig of 4112 bp in length  
\* 49988 50087: gap of 100 bp  
\* 50088 53225: contig of 3138 bp in length  
\* 53226 53325: gap of 100 bp  
\* 53326 56622: contig of 3297 bp in length  
\* 56623 56722: gap of 100 bp  
\* 56723 59393: contig of 2671 bp in length  
\* 59394 59493: gap of 100 bp  
\* 59494 63369: contig of 3876 bp in length  
\* 63370 63469: gap of 100 bp  
\* 63470 68148: contig of 4679 bp in length  
\* 68149 68248: gap of 100 bp  
\* 68249 72051: contig of 3803 bp in length  
\* 72052 72151: gap of 100 bp  
\* 72152 78079: contig of 5928 bp in length  
\* 78080 78179: gap of 100 bp  
\* 78180 82389: contig of 4210 bp in length  
\* 82390 82489: gap of 100 bp  
\* 82490 87187: contig of 4698 bp in length  
\* 87188 87287: gap of 100 bp  
\* 87288 93791: contig of 6504 bp in length  
\* 93792 93891: gap of 100 bp  
\* 93892 99177: contig of 5286 bp in length  
\* 99178 99277: gap of 100 bp  
\* 99278 104756: contig of 5479 bp in length  
\* 104757 104856: gap of 100 bp  
\* 104857 112844: contig of 7788 bp in length  
\* 112645 112744: gap of 100 bp  
\* 112745 119589: contig of 6845 bp in length  
\* 119590 119689: gap of 100 bp  
\* 119690 130279: contig of 10590 bp in length  
\* 130280 130379: gap of 100 bp  
\* 130380 144451: contig of 14072 bp in length  
\* 144452 144551: gap of 100 bp  
\* 144552 160962: contig of 16411 bp in length.

FEATURES

Source

1. .160962  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone\_lib="RPCI-11 Human Male BAC"  
1. .1187  
/note="assembly\_fragment"  
1288. .2962  
/note="assembly\_fragment"  
3063. .4458  
/note="assembly\_fragment"  
4559. .5583  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
5684. .7420  
/note="assembly\_fragment"  
7521. .9277  
/note="assembly\_fragment"  
9378. .11138  
/note="assembly\_fragment"

misc\_feature 11239. .13265  
/note="assembly\_fragment"  
misc\_feature 13366. .15440  
/note="assembly\_fragment"  
misc\_feature 15541. .17308  
/note="assembly\_fragment"  
misc\_feature 17409. .19894  
/note="assembly\_fragment"  
misc\_feature 19995. .21852  
/note="assembly\_fragment"  
misc\_feature 21953. .24275  
/note="assembly\_fragment"  
misc\_feature 24376. .27231  
/note="assembly\_fragment"  
misc\_feature 27332. .31148  
/note="assembly\_fragment"  
misc\_feature 31249. .33450  
/note="assembly\_fragment"  
misc\_feature 33551. .36754  
/note="assembly\_fragment"  
misc\_feature 36855. .39988  
/note="assembly\_fragment"  
misc\_feature 40089. .42367  
/note="assembly\_fragment"  
misc\_feature 42468. .45775  
/note="assembly\_fragment"  
misc\_feature 45876. .49987  
/note="assembly\_fragment"  
misc\_feature 50088. .53225  
/note="assembly\_fragment"  
misc\_feature 53326. .56622  
/note="assembly\_fragment"  
misc\_feature 56723. .59393  
/note="assembly\_fragment"

alignment\_scores:

Quality: 228.00 Length: 58  
Ratio: 4.145 Gaps: 1  
Percent Similarity: 94.828 Percent Identity: 79.310

alignment\_block:

US-09-512-581-2 x AC022862/rev ..

Align seg 1/1 to reverse of: AC022862 from: 1 to: 160962

351 LeuThrGluTyrLeuLysValArgSerHisaspProGluGluAlaIleAr 367  
:::|||||  
123366 ATTACAGATATTTAAAGGTTAGATCATGATCCAGGAAGCTATTTCG 123317

367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384  
|||||  
123316 TCATGATGTCTATTGTTACTATAATAACAGCTGCCAAGAGGACCTGGCCT 123267

384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400  
|||||  
123266 TAGTAATGATCAGCTGCTTGGCTTTGTAGGGAAGAACTGGATAAA 123217

401 ArgTrpArgValArgLysGluAla 408  
|||||

123216 .....CGGGTAAATCTGAGGCA 123199

seq\_name: gb\_in3:CELF39C12

seq\_documentation\_block:

LOCUS CELF39C12 40233 bp DNA INV 01-JAN-1998  
DEFINITION Caenorhabditis elegans cosmid F39C12.  
ACCESSION AF039043  
VERSION AF039043.1 GI:2736398  
KEYWORDS  
SOURCE Caenorhabditis elegans strain=Bristol N2.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.



```
10720 AGAAGAGCTAAAGAGAGT.....GTTAAGG 10695
1172 lYargLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThr 1188
:: ||| :::::||||:||||
10694 CTTACAAAGACTCGCTATCAAGAGCTAGGAATGAAAGAG..... 10654
1189 MetSerProLeuProGlyLysSerAspLysArgAspSerAs 1205
10654 ..... 10654
1205 pLeuValArgSerGluLeuGluLys.....ProArgGlyArgLysL 1219
10653 .....AAACAGAGATCGGAGAACTACTACCCCTGAGGAGGAGAA. 10612
1219 yThrProValThrGluGlnGluGluLysLeuGlyMetAsp..... 1232
:: ||||| ||| :::::
10611 .....CTATTAGAGCAAGAGATTAAAGAGCGTTAAGGCTTATTTA 10570
1233 AspLeuThrLysLeuValGlnGlnGluLysProLysGlySerGlnArgSe 1249
|| ::::: :::::|||| ||| :::::
10569 GACTGTGTATCAAGAGCTAGGAATGAAAGAGAGAAACAAGATGCGAGAA 10520
1249 rArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTTPProG 1266
10519 ATTACTACCCCTGAGCGAGGAGAAATTTTAGAGAAACAGCCCAACAA 10470
1266 luGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsn 1282
:::|||| ||| ||||| ||| |||||
10469 AAGATAAGCGATAAG...GATTGCTTGAAGAAACGCCGATCTCAACGAC 10423
1283 SerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGl 1299
||||| ||| ||| ||
10422 .....AGAGCGGCTATCATGAGTGTTCGATGG 10394
1299 yGlyThrProLysGluGluProThrMetLys.ThrSerLysLysGlySer 1315
||| ::::: ||||| |||||
10393 TTTGACCGATGAGAGAAAG.....CTCAATACCTGCGAGAGCCAGAG 10350
1316 Lys.....LysLysSerGlyProProAlaProG 1325
||| ||||| |||||
10349 AAAAGGCTGCTTGGATTCTTTGAAAACGCTAGGACCG..... 10311
1325 uGluGluGluGluGluGluGlnSerGlyAsnThrGlu..... 1338
:::||||| ||||| :::::
10310 .....ATGAAGAAAAGGAAATGCCAAACCTTTATAG 10277
1339 .....GlnLysSerLysSerLysGlnHisArgValSerArg 1350
:::||||| ||||| |||||
10276 CGATTTGATCCAGAAATCCAAATAAAGGCGACAGACAAACAA 10230
seq_name: gb_htg8:AC022862
```

```
seq_documentation_block:
LOCUS AC022862 160962 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-343C9 map 4, WORKING DRAFT
SEQUENCE, 38 unordered pieces.
ACCESSION AC022862
VERSION AC022862.3 GI:7249278
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160962)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-343C9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160962)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
```

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

## TITLE

JOURNAL

## COMMENT

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2000 this sequence version replaced gi:6978239.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6193

----- Summary Statistics

Sequencing vector: M13: M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145534 bases at least Q40

Consensus quality: 151792 bases at least Q30

Consensus quality: 154567 bases at least Q20

Insert size: 194000; agarose-fp

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1187: contig of 1187 bp in length  
\* 1188 1287: gap of 100 bp  
\* 1288 2962: contig of 1675 bp in length  
\* 2963 3062: gap of 100 bp  
\* 3063 4458: contig of 1396 bp in length  
\* 4459 4558: gap of 100 bp  
\* 4559 5583: contig of 1025 bp in length  
\* 5584 5683: gap of 100 bp  
\* 5684 7420: contig of 1737 bp in length  
\* 7421 7520: gap of 100 bp  
\* 7521 9277: contig of 1757 bp in length  
\* 9278 9377: gap of 100 bp  
\* 9378 11138: contig of 1761 bp in length  
\* 11139 11238: gap of 100 bp  
\* 11239 13265: contig of 2027 bp in length  
\* 13266 13365: gap of 100 bp  
\* 13366 15440: contig of 2075 bp in length  
\* 15441 15540: gap of 100 bp  
\* 15541 17308: contig of 1768 bp in length  
\* 17309 17408: gap of 100 bp  
\* 17409 19894: contig of 2486 bp in length  
\* 19895 19994: gap of 100 bp  
\* 19995 21852: contig of 1858 bp in length  
\* 21853 21952: gap of 100 bp  
\* 21953 24275: contig of 2323 bp in length  
\* 24276 24375: gap of 100 bp



```

11551 ATCAAGAGCTAGGAATGAAAGAGAGAAAACAAGAATGCGAGAAATTACTCA 11502
      ::: ||| |||::: ||||| ::::
920  InValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
      |||:::|||||
11501 CGCCTGAAGCGGAAAACCTA.....CysLeu.....Vallys 964
      |||:::|||||
937  LeuGluTyrMetAlaIleCysAlaLeu..... 945
      ||||| ::::||||| ::::|||||
11481 TTAGAGAAATCAAGCGCTAGATTGTTTGAAACAGCTAAACCGAAGCTGA 11432
      ||||| ::::||||| ::::|||||
946  .....CysAlaLysAspProValLys.....GluArgArgAlaH 957
      |||||:||||| ||| ::::|||||
11431 GAAAAAAGGTTGTCGAAGATCTCCCTAAAGACTTACAGAAAAGGTTT 11382
      |||:||||| ||| ::::|||||
957  IsAlaArgGln.....CysLeu.....Vallys 964
      |||:||||| ||| ::::|||||
11381 TAGCCAAAGAGAGCTGTAGGGTTTATTGGATTTCGGTTATCAAAAGCCAAA 11332
      |||:||||| ||| ::::|||||
965  AsnIleAsnValArgGluTyrLeuLysGlnHisAlaAlaValSerG1 981
      |||:||||| ||| ::::|||||
11331 AACGAAGCTGAAGAAAGAATGCCGAGAATTGCTCACCCCTGAAGCGAG 11282
      |||:||||| ||| ::::|||||
981  uLysLeuLeuSerLeuLeuProGluTyrValProTyrThrIleHisL 998
      |||:|||||: ||| ||| |||
11281 AAAGCTATTAGAAAGAGCTAAAGAGAGTGTTAAGCTTACAAA...GACT 11235
      |||:||||| ||| ::::|||||
998  euLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeu 1014
      ::::|||||: ||| ::::|||||
11234 CGGTATCAAGAGCTAGGAATGAAAAGAGAAACAAGAATGCGAGAAATTA 11185
      ::::|||||: ||| ::::|||||
1015  .....LysAspValLysGluCysLe 1021
      ::::|||||: ||| ::::|||||
11184 CTCACCCCTGAAGCTAGAAACTATTAGAGCAGAAGTTTAAAGAGCGGT 11135
      ::::|||||: ||| ::::|||||
1021  uTrpPheValLeuGlu...IleLeuMetAlaLysAsnGluAsnAsnSerH 1037
      :|||::: |||: |||:||||| |||:|||||
11134 TAAGGCTTATTAGACTCGGTTTCAAGAGCTAGGAATGAAAGAGAGAAAC 11085
      :|||:||||| ||| ::::|||||
1037  IsAlaPhe.....IleArgLysMetValGluAsn 1046
      :|||:||||| ||| ::::|||||
11084 AAGAATGCGAGAAATTACTACGCTGAAGCGGAGAAATTTTAGAGAAAT 11035
      :|||:||||| ||| ::::|||||
1047  .....IleLysGlnThrLysAspAlaGlnGlyProAspAs 1058
      :|||:||||| ||| ::::|||||
11034 CAAGCGCTAGATTGTTTGAAAACGCTAAACCGAGCT..... 10996
      :|||:||||| ||| ::::|||||
1058  pAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnI 1075
      ||||| ||| |||:|||||
10995  .....GAGAAAAAAGGTGTGTCAAGATCTCCCTAAAGACT 10959
      :|||:||||| ||| ::::|||||
1075  leIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPro 1091
      :|||:||||| ||| ::::|||||
10958 TCGCAAAAAGGTTTTCAGCTAAAGAGACGGTTAAGGCTTTATTAGACTGC 10909
      :|||:||||| ||| ::::|||||
1092  ValLeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnTh 1108
      ||| ||||| |||:|||||
10908 GTATCAAGAGCTAGG.....AATGAAAAGAGAAAAAAGAATG 10871
      :|||:||||| ||| ::::|||||
1108  rLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysP 1125
      :|||:||||| ||| ::::|||||
10870 CGAGAAGTTGCTCAGCGCTGAAGCGAGAAAACTATTAGAAGATCTAAAA 10821
      :|||:||||| ||| ::::|||||
1125  roLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAla 1141
      :|||:||||| ||| ::::|||||
10820 AAAGGCTTAAGGCTTATTGGATTTCGATTCAAAAGCCAAAACAGAGCT 10771
      :|||:||||| ||| ::::|||||
1142  GlyLysGlnSerGlnThrLysSerSerArgMetGluThr.....Va 1155
      :|||:||||| ||| ::::|||||
10770 GAAAAAAGAAGATGCGAGAAATTGCTCACCTGAAGCGAGAAAGCTATT 10721
      :|||:||||| ||| ::::|||||
1155  lSerAsnAlaSerSerSerAsnProSerSerProGlyArgIleLysG 1172
      :|||:||||| ||| ::::|||||

```



```

13428 AATATGATCGAT..... 13417
93  oGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheI 110
   |||:|||||:|
13416 .....AAGAATCTTTCTTC.... 13402
110  leThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
   :|||:|||||:|
13401 .....ANTGACGATCCCAATAGAACC 13381
127  ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
   ||| ||||| ||||| |||||
13380 TTATACAACATTTG.....AATATTGCA..... 13357
143  eCysPheGluLeuGluAspSerAsnGlnIlePheThrGlnLeuTyrArgT 160
   |||:|||||:|
13356 .....GAAATTGAGCACAACCCATTG.....AGAG 13329
160  hrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHis 176
   :|||:|||||:|
13328 CCTTTTATGAATGATATAGTAATGTTGGCAACTATGAAGAATGTTTGAAG 13279
177  MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGI 193
   :|||:|||||:|
13278 CTTATC..... 13273
193  nGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnL 210
   |||:|||||:|
13272 .....AAGACACAAAAAAC 13260
210  euAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAla 226
   |||:|||||:|
13259 TTCAAGATCAG.....ATGAAAAAGACTCTA 13234
227  GlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMetLe 243
   :|||:|||||:|
13233 GAGGCT.....TATATGACTGCATCAAAA 13208
243  uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleL 260
   :|||:|||||:|
13207 TGCCAAAACCTGAAGAAGAAAGGATCAAG.....TGCTTTAGATTTAATCA 13164
260  euGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276
   |||:|||||:|
13163 AA.....GATGAAACCTGAAAAAAGCTTACTG..... 13135
277  LeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVa 293
   |||:|||||:|
13134 .....ACCCACAAAAAGTTCAAGTGGC 13112
293  lLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerG 310
   :|||:|||||:|
13111 T.....CTAGATTGTTTGAAAAACGCTAAACCCGATGAA..... 13078
310  lnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
13078 ..... 13078
327  ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAs 343
   ||| ||||| ||||| |||||
13077 .....GAACGGAAAGAGTCCCTAANA.....CTCATAAA 13049
343  nHisPro.....AspLeuAlaLysA 350
   ||| ||||| ||||| |||||
13048 TGACCCCTGAGATTAGAGAAAAATTTCGTAAGGAATTTAGAGCTTCAAAAAG 12999
350  spleuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
   :|||:|||||:|
12998 AGCTTCAAGAGTAT..... 12985
367  ArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLe 383
   ||| ||||| ||||| |||||
12984 ...AAGATTGTATCAAAACGCAACAGAGCTGAGAAAAAC..... 12943

```

```

383  uLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspL 400
12943 ..... 12943
400  ysArgTrpArgValArgLysGluAlaMetMetClyLeuAlaGlnIleTyr 416
   |||:|||||:|
12942 .....GAATGCTTGAAGCGTTGCT..... 12922
417  LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnI 433
   |||:|||||:|
12921 .....AAGAAGCCCATAGAAGA... 12904
433  eAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleA 450
   :|||:|||||:|
12903 .....TTGAAACAACAAGCGCTAGATTGTTTGAACCAACGCTAAACCG 12861
450  sPaspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 466
   |||:|||||:|
12860 ATGAA.....GAACGGAAGAGCTGTAAATAATATTCCTCCAA 12823
467  AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaTh 483
   :|||:|||||:|
12822 GACTTGCAAAAAGAA.....CTACTAGCTGA 12797
483  rLeuAspLeuAsnAlaValLys.....AlaLeuAsnGluM 495
   :|||:|||||:|
12796 TATGAGCGTTAAGCGTTACAAAGACTGCGTATCAAGAGCTAGGAATGAA 12747
495  etrTrpLys.....CysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
   ||| ||||| ||||| |||||
12746 AAGAGAAACAAGAATCGAGAAATTAATCAAGCGCTGAAGCGAAAAAACA 12697
510 .....LeuAspLeuIleLysGlnProLysThrAspAlase 521
   ||| ||||| ||||| |||||
12696 TTAGAGAANTCAAGCGCTAGATTGTTTGAACCAAGCTAAGACCGATCAAGA 12647
521  rValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProA 538
   :|||:|||||:|
12646 ACGMAAGAGTGCTTG.....AAAAATCTCCCTA 12618
538  sProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlu 554
   :|||:|||||:|
12617 AAGACTTACAAAGCGTATATTTAGCTAAGAGAGAGCGCTGAAGCGTTATAA 12568
555  Asp.....AspGluLysIleArgLysGlnLeuGluVa 565
   ||| ||||| ||||| |||||
12567 GACTGCGCATCTCAAGCCAAACTGAAGCTGAAAAAAGAAATGCGAGAA 12518
565  lLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgG 582
   |||:|||||:|
12517 ATTACTCAGCGCT.....GAAGCGAAAAAATTTTAGAAG 12483
582  luIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeu 598
   |||:|||||:|
12482 AAGAAGCCAAA..... 12472
599  GluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspTh 615
   ||| ||||| ||||| |||||
12471 GAGAGCGTTAAGCGTTATTTGGATTGCGTATCTCAGGCCCAAACTGAAGC 12422
615  rGlu..... 616
12421 TGAACAAAAAAGAAATGCGAGAAATTTGCTACACCTGAAGCGAAAAAAGT 12372
617 .....SerIleSerAlaLeuIleLysGlnValAsnLys 627
12371 TAGACAGACCCAAAGAGCGGTAGAGCTATTGATTGCGGTATCAAAA 12322
628  SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGI 644
   :|||:|||||:|
12321 GCCAAAAACGAGCTGAAAGAAAGAAATCGAGAAATTAATCTACAGCGCTGA 12272

```











```

10100 ACTTTTAGAA.....GAAGCTAAAAAAGCGTTA 10072
671 luSerLeuLeuAlaCysLeu.....LysMetAspAspGluLys... 683
: : : : : ||| ||| : : : : : ||| ||| ||| ||| |||
10071 AGCCTATTGGATTGGTATCTCAAGCCAAACTGAAGATGAGAAAAA 10022
684 .....ValalaGluAl 687
10021 GAATGCGAGAAATTACTACCCCTGAAGCGAAAAAACTTTTAGAGCAACA 9972
687 aAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu..... 700
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9971 AGCGCTAGATTGTTGTAACCAACCTAAACGAAGCTGATAAAAAAGGT 9922
701 .....GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9921 GTGTCAAGATCTCCCT.....AAGACTTACAGAAAAAGGTTTAGCT 9878
716 HisLysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCy 732
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9877 AAAAGAGC.....GTCAAGCGCTTATTTAGACTG 9849
732 sIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheG 749
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9848 CGTATCAAGAGCTAGGAATGAAAAAGAGAAAAA..... 9815
749 luProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThr 765
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9814 .....GAATGCGAGAAATTACTACC 9794
766 Pro.....LeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9793 CCTGAAGCGAAAAAACTTTAGAAGAGCTAAAGAGAGCTGAAGCT... 9746
778 oAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValL 795
9745 .....TATA 9742
795 ysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9741 AAGACTGCTCTCTCAAGCTAGAAATGAAGAAGAAGAGGAGA..... 9701
812 TrpValProAspGluGluValSerProGluThrMetValLysIleGlnAl 828
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9700 ...GCTTGCAGAAATTACTACCCCTGAAGCGAGAAACTATTAGAGGA 9654
828 aIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnHisSerL 845
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9653 AGCTAAAAAAGCGTTAAAGCTTATTGGATTGCTATCTCAA..... 9611
845 ysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAsp 861
9610 .....GCCAAAAACGAA 9599
862 GlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLe 878
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9598 GCTGAAGAAAAAGAAATCGGAGAAATTACTCACCCCTGAAGCGAAAAAACT 9549
878 uArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCyst 895
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9548 TTTAGAGAAGAGCTTAAGAGCGCTT.....AAGCTTATAAAGACTGCC 9505
895 yrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIle 911
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9504 TCCTCAAGCTAGAAATGAACCTGAAGAGAGAGCTTCGGAGAAATTACTC 9455
912 AsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLy 928
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9454 ACCCTCAA.....GCAGAAAACTATTAGAGCAAGAAGCTTAAAAA 9414
928 sGly.....LeuSerArgLeuArgLeuProLeuG 938
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9413 GAGCGTCAAGGCTTATTTAGACTGCGTTTCAAGAGCTAGGAATGAAAAAG 9364

```

```

938 luTyrMetalIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
|| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9363 AGAAACAAGAAATCGAGAAATTTACTCAGCCTGAAGCAAGAAATTTTA 9314
955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgG1 971
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9313 GAGAAGCAAGCATTAAGTTGTTGGAAAA.....GCTAGAAATGA 9273
971 uTyrLeuLysGlnHisAlaValSerGluLysLeuLeuSerLeuLeuP 988
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9272 A.....GAAGAAAGAAAGCATGTTTAAAAATCTCTTAAAGATTTC 9229
988 roGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9228 AGAAAAATGTTTAGCTAAA..... 9209
1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLe 1021
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9208 .....GAGAGCCTTAAAGCTTATAAAGACTGCCT 9180
1021 uTyrPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHisA 1038
|| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9179 C.....TCTCAAGCTAGAAATGAAGAAAGAAAGAGAG 9148
1038 laPhe.....IleArgLysMetValGluAsnIle 1047
|| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9147 CTTCGAGAAACTACTCACCCCTGAAGCGAGAAACTCTTAGAG..... 9104
1048 LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLy 1064
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9103 .....CAAGAAGTTAAGAAGAGCGTTAA 9081
1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSert 1081
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9080 GGCTTATTGGATTGCGTTTCAAGAGCTAGGAATGAAAAAGAAACAGG 9031
1081 hrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9030 AATGCGAGAAATTCCTCACCCCTGAAGCGAGAAACTTTTAGAGAAAGAA 8981
1098 PheThrGlnProAspLysAsnPheSerAsn...ThrLysAsnTyrLeuPr 1113
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
9890 CTCACAAAAAGATAAAGCGATAAAGATTGCTTAAAAAACCGCGATCC 8931
1113 O.....ProG 1115
9890 TAACGACAGAGCGCTATCATGAAGTGTTCGGATGTTGAGCGATGAAG 8881
1115 luMetLysSerPheThrProGlyLysProLysThrThrAsnValLeu 1131
|| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8880 AAAAGCTTAATATTTCAGAGAGCTAGAGAAAGGCTGTTCGGATTGT 8831
1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8830 TTGCTATGCTAAACCGATGAAGAAAAAGGAATGCCAAAACCTTTA 8781
1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProS 1165
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8780 TAGTGATTTCACAGAAATCCAAAATAAAGGGCACAGAACAAACAAA 8731
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGlu...MetAsp 1180
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8730 ATCAATGTAGTAAACAGAAAGGTTGCATCAAGCAAGCGAGTCTTGAT 8681
1181 HisSerGluAsn.....GluAspTyrThrMetSerSerProLeuProG1 1195
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8680 AACTTAGTACCTACTGATCAACAAGCCATAGAGCAATGTTTAGAGGG 8631
1195 yLysLysSerAspLysArg.....A 1202
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
8630 CTTGAGCGTAGTGAAGGCGCTAATTCCTAGGAATTAACGACAGACTG 8581

```



```

142 nLeuCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrA 159
|||||:|||||:|||||:
11629 .....GAAATGAGCAAAACCATTTG.....A 11605

159 rgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMet 175
|||||:|||||:|||||:|||||:
11604 GAGCCTTTATGAGTGTATGTTGTAATGGTGGCACTATCAAGAATGTTTG 11555

176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSe 192
|||||:
11554 AAGCTTATC..... 11546

192 rgLlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysA 209
|||||:
11545 .....AAGACAAAA 11536

209 snLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThr 225
|||||:|||||:
11535 AACTTCAAGATCAA.....ATGAAAAGACT 11510

226 AlaGlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMe 242
|||||:|||||:
11509 CTAGAGGCT.....TATAACGACTGCATCAA 11484

242 tleuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuI 259
|||||:|||||:
11483 AAATGCTAAACTGAAGAAGAGGATCAAG.....TGTTAGATTTAA 11440

259 leuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
|||:|||||:|||||:
11439 TCAAA.....GATGAACCTGAAAAAGCTTACTGAAC 11405

276 Gln.....LeuGluPheLysLeuLysSerAsnAspAs 286
|||||:
11404 CAACAAAAAGTTCAAGTGGCGCTAGATTGTTGAAAAACGCTAAAAACCGA 11355

286 nGluGluArgLeuGlnValValLysLeuLeu..... 296

11354 TGAAGAACGAAACGAGTGTCTGAAACTCATNAATGACCCTGAGATTAGAG 11305

297 .....AlaLysMetPheGly..... 301

11304 AGAAATTCGTAAGGAATTAGGCTTCAAAAAGAGCTTCAAGAGTATAAG 11255

302 .....AlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrG1 316
|||||:|||||:
11254 GATTGTATCAAAAACGCCAAAACAGAGGCTGAGAAAAAT.....GA 11214

316 nCysTyrLeuGlyArgPheAsnAspIleHisValProIleArg..... 330

11213 ATGCTTGAAGGCTTGTCTAAAGAAGCTATAGAAAAGATTGAAACAGCAAG 11164

331 ..LeuGluCysValLysPheAla.....SerHisCys 340
|||||:|||||:
11163 CGCTAGATTGTTGAAAAACGCTAAAAACCGATGAAGAACGAAACGAGTGC 11114

341 LeuMetAsnHisPro...AspLeuAlaLysAspLeuThrGluTyrLeuLys 356
|||||:|||||:
11113 TTGAAAAATATTCCCAAGACTTGGCAAAAAGAACTACTAGCTGATATGAG 11064

356 sValArgSerHisAspPro.....GluGluAlaIleA 367
|||||:|||||:
11063 CGTCAAGGCTTATAGGACTGCGTTTCAAAAGCTAGGAATGAAAAGAGA 11014

367 rgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeu 383
|||:|||||:
11013 AAAAAGATGCCAGAAATGCTCACGCCCTGAAGCGGAGGAAAGTTA... 10967

384 LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
|||||:|||||:
10966 .....GAACAACAGGTTCTAGATTGTTTGA AAAACGCTAAAAACCGATGA 10923

400 sArgTyrArgValArgLysGluAlaMetMetGlyLeu..... 412

```

```

10922 AGAA.....CGAAAAAAGTGTTTGAAAGATCTCCCTAAAGACTTAC 10882
|||||:|||||:|||||:
413 ..AlaGlnIleTyrLysTyrAlaLeuGlnSer..... 423
|||||:|||||:
10881 AAGCGATATTTAGCTAAAGAGAGCGCTGAAAGCTTATAAGATTCGGTA 10832

424 .....AlaAlaGlyLysAspAlaAlaLysGlnIleAl 434
|||||:|||||:
10831 TCTCAAGCCAAACCGAAGCTGAGAAAAAAGAAATGCGAGAAATTCGCTAC 10782

434 atrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp. 450
|||||:|||||:
10781 CCTCAAGCGAAAAAAGCTTTTGAAGAAGAACCAAGAGAGCGGTTAAAG 10732

451 .....AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVal 464
|||||:|||||:
10731 CTTATTTGGATTGCGTATCTCAAGCCAAAGCTGAAGCTGAGAAAAAGAA 10682

465 ProHisAsnLeuGluThrThrGluArgMetLysCysLeu..... 477
|||||:|||||:
10681 TCGAGAAATTTACTCACCCCTGAAGCGAAAAAAGTTAGAAAGAGCTAA 10632

478 .TyrTyrLeuTyrAlaThrLeuAsp...LeuAsnAlaValLysAlaLeuA 493
|||||:|||||:
10631 AAAAAAGCGTTAAGGCTTATTTGGATTGCGTATCTCAAGCCAAAGCTGAAG 10582

493 snGluMetTrpLysCysGlnAsnLeuArgHisGlnValLysAspLeu 509
|||||:|||||:
10581 ATGAGAAAAAGAAATGCGAGAAATTTACTCACCCCTGAGCGGAAAAAAT 10532

510 .....LeuAspLeuIleLysGlnProLysThrAspAlaSe 521
|||||:|||||:
10531 TTAGACAAACAAAGCGCTAGATTGTTGAAAAACGCTAAAAACCGAAGCTGA 10482

521 rValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProA 538
|||||:|||||:
10481 TAAAAAAGG.....TGTTGTCAAAGATCTCCCTA 10453

538 spProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlu 554
|||||:|||||:
10452 AAGACTTGCAGAAAAAGTTTGTAGCCAAAGAGAGCGTTAGGTTTATTWG 10403

555 Asp.....AspGluLysIleArgLysGlnLeuGluVa 565
|||||:|||||:
10402 GATTGCGTATCAAGACCCAAACGAAAGCTGAAGAAAAAGAAATGCGAGAA 10353

565 lLeuValSerProThrCys.....SerCysLysGlnAlaG 577
|||||:|||||:
10352 ATTACTCACCCCTGAAGCGGAAAAACTATTAGAAGAGCTAAAAAAGCG 10303

577 luGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPro 593
|||||:|||||:
10302 TTAAAGCTTATAAAGACTGCGTGTCAAGACCGAGGAATCAAAAAAGAGAA 10253

594 ThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPr 610
|||||:|||||:
10252 AAA.....GAATGCGAGAAATTTACTC..... 10232

610 oValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnL 627
|||||:|||||:
10231 .....ACCCCTGAAGCGGAAAAACTATTAGAAGAGCTAAAAA 10195

627 ysSerIleAspGlyThrAlaAsp.....AspGluAsp 637
|||||:|||||:
10194 AAGCGTTAAAGCTTATTTGGATTGCGTATCTCAAGCCAAAGCTGAAGAT 10145

638 GluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLy 654
|||||:|||||:
10144 GAGAAAAAAGAAATGCGAGAAATTTACTCACCCCT.....GAAGCGAAAAA 10101

654 sValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheG 671
|||||:|||||:

```



```

gene
CDS
    complement(3357..5603)
    /gene="cag-beta"
    complement(3357..5603)
    /gene="cag-beta"
    /note="HP524; cag5; JHP473; ORF10; similar to VirD4"
    /codon_start=1
    /transl_table=11
    /product="Cag-beta"
    /protein_id="AAF80195.1"
    /db_xref="GI:8843974"
    /translation="MEDLEYNTLYFIEDYKLVVIPSFTGLIALFFLYFIKAQKAFK
    DRANQPKQKFEIIDDLEKRVKTFQWLAQLLLSYFISGLFVILILGNFYDD
    KNSPESDDDLFWIAYATQDPNRYFKALGFSKLIYGFNLSLVYSSILCSYIFITF
    FWFVLYLTRTRDYGANKVDLFGSASWETEEMKAKLITPNKKRAFDKREIVIG
    RRGLDFTAYAGQAFGLIAPTRSGKGVGFIMPNNIYPONLIVFDPKADWETCGKI
    REKRNQKVIYEFSLTKRHPNFAYVDFGVVLTEDILSOLITRLKHGHWASGG
    DFTQIFGLAKLVFPERPNEKDPFNSQARNLVINCNIYRDLMMTKKGLFVVRKKI
    IMPDPTMFFTGSMASGILNIDEDTNMEKVVSLEMFEGGEDKSGDNLRLASPATRNM
    WNFKTMGGARETYSVGVYTSAPAPYNNAMIRNFTSANDPFRRLRIDEVSIGVIA
    NPKESTIVGPILFELFNNVIYNSLILPHDPOCKRSCMLMDEFTLCGYLTFYKAVG
    IMAEVNMPAFVFSKAOLENDPPLGYGRGAKTILDLNLSNMYGINNNDNYKHEFEK
    LSKHALYTRDVSRIIDNTCKNTSISNKRFLMTPEDELTWGDDELILILENTLKPI
    KCHKALYDDPFFDELILKVSPLSKYKGLKVPDQTFTFYDDLQAAKTRGELSVDKSL
    VPVGSSEL"
    complement(5612..6604)
    /gene="cag-alpha"
    complement(5612..6604)
    /gene="cag-alpha"
    /note="similar to VirB11; HP525; JHP474, ORF11"
    /codon_start=1
    /transl_table=11
    /product="Cag-alpha"
    /protein_id="AAF80196.1"
    /db_xref="GI:8843975"
    /translation="MTEDRLSAEDKKFLEVERALKEAALNPLRHATELFGDFLKMEN
    ITECYNGKVVYLVKNNGEWPFDVRDKAFSLSRMHFARCCASFKKKTIIDNYENP
    ILSNLNGLANGRVQIVLSTPVNDDETISIRIPSKTTPVPHSFEEQGYFNLDNKEQA
    ISAKIDGAIKNTVCGGTGKTYIKSIIMEFIPKEERIISIEDTEIYFVKHHNY
    TQLFEGGNTSADCLSKCLMRPDRILILGELRSSEAYDFYNVLCSGHGKTLTLHAGS
    SEEFIRLANNSSNSAARNIFESLIEFGKLDIMVHINHKKQCDDEFYIKHR"
    complement(6609..7208)
    /gene="cag-z"
    complement(6609..7208)
    /gene="cag-z"
    /note="HP526; cag6; JHP475; ORF12"
    /codon_start=1
    /transl_table=11
    /product="Cag-z"
    /protein_id="AAF80197.1"
    /db_xref="GI:8843976"
    /translation="MELGFNEARQKILDSNLSLRNANVYRDFEIQNYATSLKDSND
    PQDFURRVOELRINKQNFISFDAYINLNLVLASYNRKQKERTFAESTIKNELTLG
    EFVAFISDFNFTNDEVARISDLVASYLPREYLPFPFDGNNMGVAFQILGIDDFGKK
    LNEIVQDICTYKIIILSKNKTLYLSLERAKLITQLKLNLE"
    complement(7346..12517)
    /gene="cag-y"
    complement(7346..12517)
    /gene="cag-y"
    /note="similar to VirB10; HP527; cag7; JHP476; ORF13 and
    ORF14"
    /codon_start=1
    /transl_table=11
    /product="Cag-y"
    /protein_id="AAF80198.1"
    /db_xref="GI:8843977"
    /translation="MNEENDKLETSSKTIQODSPQDLSNEEATEANHFEDLLKESSDN
    HLDNSAETKSNFEDKLETOITOMDSGGNETSESSNGSLADKLFKKARKLVYDKRPPT
    QOKSDEEAOKLNEDDOENNEYQETOTGLIDDETSKKAQODSPQDSNNEATEANH
    PEDLLKRESSNHLNDSATKSNFEDKLEETDSDNSDQEIIGSKKKYIIGGIVAV
    LILVLFSSRFHYFMPYLEDKSSRFKORNLVDEIQIRQEIYNLLKREKERNMND
    KNLVFDPPNTLFLYNLIAIEDKNPLRAFYEYICNGNGNVEECLKIKDKLOQMD
    KTLFAYNDICIKNAKTEERIKCLDIKDENLKLKSLNQQKQVQVLDLCKNAKTERN
    ECLKLIINDPEIRKPRKELGLQKLEQYKDCIKNAKTEAKNECLKGLSKEAIERLKQ

```

```

alignment_scores
    Quality: 229.00      Length: 1586
    Ratio: 0.331        Gaps: 69
    Percent Similarity: 43.632    Percent Identity: 17.844

alignment_block
    US-09-512-581-2 x AF282852/rev ..

    Align seg 1/1 to reverse of: AF282852 from: 1 to: 21347

    2 AlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProG1 18
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11986 GCAGAACTAAGACCAATTTTGTAGAGACAAGCTA..... 11951

    18 yValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgL 35
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11950 ....GAAGAAATAACTGACGACTTAACGATCAAGAGATTATCAAGGAA 11905

    35 euLys..... 36
      |||

    11904 GCAGAAAGAAATACATTATTGGTGGCATTGTAGTCGCTGTTCTTATCGTG 11855

    37 .....MetValValLysThrPheMetAspMetAspG1 47
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11854 ATTATTATTATTTCTAGAACGATTTTTCACCTACTCTCATGCTTTTGAAGA 11805

    47 nAspSer.....GluGluGluLysGluLeuTyrLeuAsnLeuAlaL 61
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11804 TAAAGCTCTCGTTTGTAGCAAGATAGGAATCTTTATGTCAATGATGAAA 11755

    61 euHisLeuAlaSerAspPhe.....PheLeuLysHisProGlyLysAsp 75
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11754 TCCAAATAAGCGCAAGAGATATAACCGATTGCTGGAAGAACGAGTGAATA 11705

    76 ValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAl 92
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11704 GGCAATATGATCGAT..... 11690

    92 aProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetP 109
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11689 .....AAAAATCTTTTCTTC. 11675

    109 heIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPhe 125
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11674 .....AATGACGATCCCAATAGA 11657

    126 AsnArgTyrPheTyrLeuLeuGluAsnIleAlaTyrValLysSerTyrAs 142
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    11656 ACCTTATACAACTATTG.....AATATGCA..... 11630

```



4318	AAACCCAAAGATTACGACTACTAAGATAGAAATAGAAATGCTCGGATT	4367
1246	rGlnArgSerArgLysArgGlyHis.....ThrAlaSerG	1258
4368	GTCCAAAGACCTTCAGAAAGTCATGATGAATGAATCTGTAGCTAAGG	4417
1258	luSerAspGluGlnGln.....TrpProGluGluLys	1268
4418	AGAAGATGACCTACAGAGGCTGCAAGAGATTCTCAATCTGAAAGTGAC	4467
1269	ArgLeuLysGluAspIle.....LeuGluAsnG	1278
4468	CAGCTCAAGAAACATAAAGAAATGTAGCTAAACACTGGAACACTGA	4517
1278	uAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPro	1294
4518	AGAGGAACCTAAAGTTGCTCATTTGTCCTGAAGAACAGAGAACTA	4567
1295	..LysProLeuGlyGlyThrProLysGluGluProThrMetLysThr	1310
4568	TTAATGAGTTAAGAGTGAATCTTTCAGAGAAGGAAACTGAAATATCAAC	4617
1311	SerLysLys.....GlySerLysLysLysSerGlyProPr	1322
4618	ATTCAAAGACAGTTAGAACATCAATGATGATAATACAGAACAGATCCA	4667
1322	oAlaProGluGluGluGluGlu.....GluArgGlnSerGlyA	1336
4668	AGAGATTATGAGAAGAGGACAACTTAATATAACAACTAGTGAGG	4717
1336	snThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla	1352
4718	TTCAGGAAACCTGTAAGTGAACCTCAACAACTCAAGGAGCATCGCAAGCC	4767
1353	GlnGlnArgAla...GluSerProGluSerSerAlaIleGluSerThr	1367
4768	AGGATTCAGCTACTCAAAAGTATAGAAAGTAAAGTCTCGAGTTGACC	4815
seq_name: gb_ba2:AF282852		
seq_documentation_block:		
LOCUS	AF282852 21347 bp DNA BCT 07-AUG-2000	
DEFINITION	Helicobacter pylori cagII pathogenicity island, complete sequence.	
ACCESSION	AF282852 U60176	
VERSION	AF282852.1 GI:8843969	
KEYWORDS		
SOURCE	Helicobacter pylori.	
ORGANISM	Helicobacter pylori.	
REFERENCE	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.	
AUTHORS	1 (bases 1 to 21347) Censini,S., Lange,C., Xiang,Z., Crabtree,J.E., Ghiara,P., Borodovsky,M., Rappuoli,R. and Covacci,A.	
TITLE	cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14648-14653 (1996)	
MEDLINE	97121442	
REFERENCE	2 (bases 1 to 21347) Covacci,A. and Rappuoli,R.	
AUTHORS	Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell	
TITLE	J. Exp. Med. 191 (4), 587-592 (2000)	
JOURNAL	J. Exp. Med. 191 (4), 587-592 (2000)	
MEDLINE	20150112	
PUBMED	10684850	
REFERENCE	3 (bases 17223 to 21347) Censini,S., Rappuoli,R., Lange,C. and Covacci,A.	
AUTHORS	Direct Submission	
TITLE	Submitted (06-JUN-1996) Molecular Biology, Chiron-Biocrine, Via	
JOURNAL	Firenze 1, Siena, SI 53100, Italy	
REFERENCE	4 (bases 1 to 21347) Censini,S., Rappuoli,R. and Covacci,A.	
AUTHORS	Direct Submission	
TITLE	Submitted (17-MAR-2000) Molecular Biology, Chiron-Biocrine, Via	
JOURNAL		
REMARK	Firenze 1, Siena, SI 53100, Italy	
FEATURES	Sequence update by submitter	
Location/Qualifiers	1..21347	
Source	/organism="Helicobacter pylori"	
	/strain="CCUG 17874; NCTC 11638"	
	/db_xref="taxon:210"	
	/note="type I-specific and disease-associated virulence factor genes; partial sequence of cag pathogenicity island presented in AF282853"	
misc_feature	312..342	
	/note="cag pathogenicity island integration signal"	
misc_feature	343..19467	
	/note="cagII pathogenicity island"	
gene	474..821	
	/gene="cag-zeta"	
CDS	474..821	
	/gene="cag-zeta"	
	/note="HP520; cagI; JHP469; ORF6"	
	/codon_start=1	
	/transl_table=11	
	/product="Cag-zeta"	
	/protein_id="AAF80191.1"	
	/db_xref="GI:8843970"	
	/translation="MADTINTTETHETKPNFVNFKNLTKDRYDSLGLGAGVLCCVLSGAMGIVGIIIFVAIGIFLSPFNINLVKLEKSKSKAAATVNNETQKSQATSVTNEPTAKETKD"	
gene	929..1279	
	/gene="cag-epsilon"	
CDS	929..1279	
	/gene="cag-epsilon"	
	/note="HP521; cag2; JHP470; ORF7"	
	/codon_start=1	
	/transl_table=11	
	/product="Cag-epsilon"	
	/protein_id="AAF80192.1"	
	/db_xref="GI:8843971"	
	/translation="MIQRELNSVFGCSYVNRGCKGVKLVREINGSKMKVCLASVKA NIKENSPFLGCKNYLKCKWKTASVDSQDPKYPKCNGLMKRKNFNNEFTATSLTKC NGILFLYQSEKGN"	
gene	1279..2724	
	/gene="cag-delta"	
CDS	1279..2724	
	/gene="cag-delta"	
	/note="HP522; cag3; JHP471; ORF8"	
	/codon_start=1	
	/transl_table=11	
	/product="Cag-delta"	
	/protein_id="AAF80193.1"	
	/db_xref="GI:8843972"	
	/translation="MFRKLATAVSLIGLLASNTLYAKEISEADKVIKATKETKTKKE AKRLKKAQRQIPDHKKPOYASVDDTKQALFDIVDTLVNNDKSGDFGNSALKDK KTYLYAMDLDYNNVLSSTENPIKTRAMGTYADLIITGSLQVNGYVNIILKLNKR N AKFLVKNENMPYAQATFLRVPKRSNDPNAHLDKGAIDENKLFQQRATFNANDV ICRPNDEVCSPLEDEMAMPYSDSKORPNIIAPYSYLRKETNNANQAQSPATQF APENSKELTEELIANSQLIANEEEREKKLAEKEQAEALAKYKLKDLNOKKLVL EAEKKNAKPRVVEVPSPQTSNDSDETRVKEKENYNGLLVDKETTIRKSYEGTL ISENSKKTPLNPNDLRSLEETIKSVYIKSNGLCYTNGINLYVKIKNDPVKEGMLCG YESVONLLSPLKDKLVYDKQKQLKALLKDSK"	
gene	2729..3238	
	/gene="cag-gamma"	
CDS	2729..3238	
	/gene="cag-gamma"	
	/note="HP523; cag4; JHP472; ORF9"	
	/codon_start=1	
	/transl_table=11	
	/product="Cag-gamma"	
	/protein_id="AAF80194.1"	
	/db_xref="GI:8843973"	
	/translation="MFGKWLGLTLFLNSLAYPCOKVITTFKQYENLTQIROKGDNEV VCPFLISLALLESLGLNNKREKSLKDTYSMFHITNTAKKFYPTYSKTLKYLKN DVDFAIQLAKLENFDYKQKYPNKSIVQLVEMAIGTYNGMKHNPNGAYVKKFRCL IYSQVRYNE"	



3541 AGTGAGATGCAGAAAAAGATTAAATGAAATAGAGAAATTTAAAGAAATGAATT 3590  
 982 LysLeuLeuSerLeuLeuProGluTyValValProTyThrIleHisL 998  
 3591 AAAGAACAAAGAATTGACATTGGACATATATGGAACATATGGAACAGAGAGCTTGAGT 3640  
 998 euLeuAlaHis...AspProAspTyrValLysValGlnAspIle...Glu 1012  
 3641 TGGCTCAGAAACTTAATCAAAATTTATGAGGAAGTGAATCTATAACCA 3690  
 1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMet 1029  
 3691 GAAGAAAGATTCTAAAGGAA...TTACAGAGATCAATT 3725  
 1029 talLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluA 1046  
 3726 TGAACAGAGAGAGACCCATTAGAGGATATTAAGAGAATTTGAAGCTA 3775  
 1046 snIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn 1062  
 3776 CAGGCTACAAACCAAGAAGCACTAAAAATTGCTCATATTCACCTAAAA 3825  
 1063 GluLysLeuTyrThrValCysAspValAlaMetAsnIleLeuMetSerLy 1079  
 3826 GAACACCAAGAACTATTGATGAATTAAGAGAAGACGCTA...TCTGA 3869  
 1079 sSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaA 1096  
 3870 GAAGACA...GCTC 3880  
 1096 rgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112  
 3881 AAATAATAAATACTCAGGACITAGAAAATCCCATACCAA...TTA 3924  
 1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAs 1129  
 3925 CAAGAAGAGATCCCA... 3939  
 1129 nValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerG 1146  
 3940 GTGCTTCATGAGCAACGAAGATTACTGCCCTAAATGTGAAAAGACTGAGT 3988  
 1146 InThrLysSer...SerArgMetGluThrValSerAsnAlaSerSer 1160  
 3989 AGACTCAGGAACATGATGAATGGAGTTATTAAACAGAACACTCCACA 4038  
 1161 SerSerAsnProSerSerProGlyLysGlyArgLeuAspSerSe 1177  
 4039 ACCAAGGACTCAACAACACTGCGCAAGATA... 4068  
 1177 rGluMetAspHisSerGlu...AsnGluAspTyrThrMetSerSerProL 1193  
 4069 GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAAAGTCCAGGAAG 4117  
 1193 euProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209  
 4118 AGATAAATCTCTAACCAAGGAAGAGACACCTTAAACCATTAAGAA 4167  
 1210 GluLeuGluLysProArgLysLysLysThrProValThrGlu... 1224  
 4168 GCCCTTGAAGTTAAACATGACCAGCTGAAGAACAATATATTAGAGAAACTTT 4217  
 1225 ...GlnGluLysLeuGlyMet. 1231  
 4218 GGCATAAATCCAGAGAGTCTCAAGCAACAAAGACAGACTCTTAATATGA 4267  
 1232 ...AspAspLeuThrLysLeuValGlnGluIn... 1241  
 4268 AGAAAAAGACAACTGAACTACCAAAATCGTGTAGTGTGAGATGGACCAATTC 4317  
 1242 LysProLys...Glyse 1246



[illegible]

449 IleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 455  
 1990 .....GCCTTGATGCCAAGAGAGAAATCAACGCTTTCTTGAAG 2027  
 465 oHisAsnLeuGluThrGluArgMetLysCysLeuTyTyLeuTyrA 482  
 2028 TGAANAATCTGGAGTTGAAGGACAAATGAAGAAGACTTGCACACTACATACA 2077  
 482 IaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498  
 2078 AGCAAAATGGAA.....AATGATATTCAGTTATAT 2106  
 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLleLysG1 515  
 2107 CAAGCCAAATTTGGAGGCAAAAGAAATGCAAGTTGATCTGGAGAAAGA 2156  
 515 nProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValI 532  
 2157 A.....TTACAATCTGCTTTTAATGAGATAACAAAC 2188  
 532 leThrArg.....AsnLeu 536  
 2189 TCACCTCCCTTATAGATGGCAAGATTCCAAAGATTTCCTCTGTAATTG 2238  
 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLe 553  
 2239 GAATTCGGAAGAAAGATTACTGATCTTCAGAAAGAACTAAATAAGAAAGT 2288  
 553 uGluAspAspGluLysIleArgTysGlnLeuGluValLeuValSerProt 570  
 2289 TCAAGAAAATGAAGCTTTGCGGGAAGAAAGTC.....ATTTTGCTTTTCA.... 2331  
 570 hrCysSerCysLysGluAlaGluCysValArgGluIleThrLysLys 586  
 2332 .....GAATTGAATCTTTACCTTCTGAAGTAGAAAGCGCTGAGGAAGAG 2376  
 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLleLysPh 603  
 2377 ATA.....CAAGACAAATCTGAAGAGCTCCATAATAACATC 2414  
 603 leuLeuGluArgIleAlaProValHisIleAspThrGluSer....IleS 619  
 2415 AGAAAAGATAAATGTGTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCT 2464  
 619 erAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635  
 2465 AAGGTTTACTTGAAGAAATTCGGGAACAAAGATGACCTAGCAACTACA 2514  
 636 GluAspGluGlyValProThrAspGlnAlaIle..... 646  
 2515 CAGTCGAATTAATAAGCACTGATCAAGAAATTCMAAATTTCAAAACCCCT 2564  
 647 .....A 647  
 2565 TCATATGCACTTCAGCAAAAGTATAGATGGCTCTTCAGCAGCAATCAGA 2614  
 647 rgAlaGlyLeuGluLeuLysValLeuSerPheThrHisProIleSer 663  
 2615 GAATGAATCAGGAAATAGTTAATCTCTAAA..... 2646  
 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAs 680  
 2647 .....GAAGCCCAAAATTTGATTCGAGTTGGTGCTTTGAAGACCGA 2690  
 680 p.....AspGluLysValAlaGluAlaAlaL 689  
 2691 GCTTTCTTACAAGACCCCAAGAACTTCAGGAGAAACACGTCGAGGTTCAAG 2740  
 689 euGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705  
 2741 AAAGACTA.....AATCAGATGCAACAGCTGGAAGCAACAAATAGAAAT 2784







```

11643 .....TCGACAGAAGTAAGTTTATT 11663
794 ValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLy 810
      ||||| ||| |||
11664 GTTAAA.....CTTACTAAC..... 11678
810 LeuTrpValProAsp..... 815
      : : : :
11679 ....TTTCATTTTGTGACTATATATGATGATGATGATGATGATGATG 11724
815 ..... 815
11725 TATTTCACAGGTTTATGAGGTGTCATATAAATCTTATTGACGTTAA 11774
816 GluGluValSerProGluThrMetValLysIleGlnAlaLysMetMe 832
      ||||| ||| |||
11775 GAAGATGCATCAAGTGAAGTAACTTCTCTATTGGTATTATTAATAATCTC 11824
832 tValArgTrpLeuLeuGlyMetLysAsnHisSerLysSerGlyThrS 849
      : : : :
11825 TATTATTCATCAACTGGTTTATATAGAAATTTTCCCA..... 11861
849 erThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThr 865
      : : : :
11862 .....CTGATGGTCTCGCAGGTACATATCTCATGCGCATGTT... 11900
866 GluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
      ||||| ||| |||
11901 GAAGAGCTTAATCTGAAGAAAGAACGCTTTTAAAGATAATCGAGGATAAATC 11950
882 aGlySerAlaIleValLysLeuLaglnGluProCysTyrHisGluIleI 899
      : : : :
11951 TTGAGCAGTGGAGGTGAAATTTCTTA.CATTCTATCATTCACCATCTT 11999
899 leThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys 915
      : : : :
12000 TATATTTACCAAAATTCAA...TGATCTGGTTTCCCTAAATAAATCTA 12046
916 TyrGlnValArgGlnValPheAlaGlnLysLeu..... 926
      : : : :
12047 AGCAGGATAAGGAAGATGATCTGCTGAGTCTACTCTTTATTCGCTTG 12096
927 .....HisLysGlyLeuSerArgLeuArgLeu..... 935
      ||||| |||||
12097 TAAGTGAAATCTCCATAGTCTATGATATACCCACAATTTATAATTTAAT 12146
936 .....ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 949
      ||||| |||||
12147 TTAGCTTTACTCTTGAGTTTGTGCTGT.....ATGTCAGT..... 12184
950 ProValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnI 966
      : : : :
12185 ...ATACAAAGGAGAAATCCAAAGAGAGAAATTTGCTCTAAGAATGT 12231
966 eAsnValArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysL 983
      : : : :
12232 GGAACCG..... 12238
983 euLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeu 999
      ||||| |||||
12239 .....AGTAGTCTCCAGAA..... 12253
1000 AlaHisAspProAspTyrValLysValGlnAspIle..... 1011
      ||||| |||||
12254 .....GTCAGGTATGAAGTATATATAAGAATTTCTAGTT 12285
1012 .....GluGlnLeu.....LysAspValLysGluCysLeuTrpP 1023
      ||||| |||||
12286 TTAGTTGTTTGAAGTTTGTCCGTGAGTGAATTTAGTTTCAATTTATGG. 12334
1023 heValLeuIleLeuMetAlaLysAsnGluAsnHisSerHisAlaPhe 1039
      : : : :
12335 .....ATGTAGATCTCTATGCAACAATGAAGAAGAA..... 12367

```

```

1040 IleArg.LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyP 1056
      ||| : : : : : ||||| ||||| : : : : : |||
12368 ...AGACTCTGTAAACAGACTCCATTAAAGCAACAAAGAACCAAGAGTG 12414
1056 roAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
      ||| |||||
12415 CACTGAAGGCTGTAAACAATGAACCA..... 12440
1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLy 1089
      : : : : : ||||| : : : : : ||||| |||
12441 .....GAAAGCACTACAGGGAATAATCTTAATCCTTGAA 12475
1089 sAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPhes 1106
      | : : : : : : : : : : ||||| ||||| : : : : :
12476 A.....AAGCTGAATGGTGAACCTGATAAAACAAGAG 12507
1106 erAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrPro 1122
      : : : : : |||||
12508 GCAGAACT..... 12515
1123 GlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSe 1139
      ||||| : : :
12516 GCACAAACACAG..... 12527
1139 rSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValS 1156
      ||| : : : : : : : : : : |||||
12528 .....AAGGTGACTCAAGCTATGCACCGGAAATTCGAAAAA.... 12563
1156 erAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGly 1172
12563 ..... 12563
1173 ArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMe 1189
      ||| : : : : : ||||| : : : : : |||||
12564 .....GATTGTGATGAGCAGGAGACCTCGAAACCAAGATGACAAGA 12607
1189 tSerSerProLeuProGlyLysLysSerAspLysArgAspSerAspL 1206
      ||| ||| ||||| : : : : : |||||
12608 CAGTCTGAAATTG...GGGAAAGAATCAGATGCA.....GAGCCTGATC 12648
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
      : : : : : ||||| : : : : : ||||| : : : : :
12649 GTATGGAAGATCACCAGAATTGCTGAAATTCACAAATGTAGAAACCAA 12698
1223 ThrGluGlnGluLysLeuGlyMetAspLeuThrLysLeuValG 1239
      ||||| |||||
12699 ACTGATGGAGAAGAG.....CA 12715
1239 nGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThra 1256
      ||||| : : : : : ||||| : : : : : |||||
12716 GGAGGCGAGCAAGAGCCCAACGCGCAGAGTCTAAACTAATGCGAGAGGC 12765
1256 laserGluSerAspGluGlnGlnTrpProGluLysArgLeuLysGlu 1272
      : : : : : ||||| : : : : : ||||| : : : : :
12766 CAAATGCAGAACCCGAACTGATGGAAGAGCATATAATCATTTGAAGGAG 12815
1273 AspIleLeuGlu.....AsnGluAspGluGlnAsnSerProProLysLy 1287
      ||| : : : : : ||||| : : : : : |||||
12816 CCAATGCAGAGCCCCAAATCTGATGGAGAAGACGAGGAGCAGCAAAAGA 12865
1287 s.....GlyLysArgGlyArgProProLysP 1296
      : : : : : ||||| : : : : : |||||
12866 GCCAAATGCTGAGTCAAAACTGATGGCAAAAATCAGGAGGAGCAAAAG 12915
1296 roLeuGlyGlyThrProLysGluGluProThrMetLysThrSer... 1311
      ||| : : : : : ||||| : : : : : |||||
12916 AGCTAACTGCAGAACCGCAAACTGATGAGGAAGAGACCAAGGTAGTGAT 12965
1312 .....LysLysGlySerLysLysSerGlyProProAlaProG 1325
      : : : : : ||||| : : : : : |||||
12966 GAGGTAGACAAAGTCACAGAAAGACAAATGTAGAACCGGAGCTGA 13015

```



[illegible][illegible]







5342 GCAGGAGCTGCAGAGCCATTAAGCAGCAGGAGCAATCTAAGCAAGAA 5391

1346 HisArgValSerArgAlaGlnGlnArgAlaGluSerProGlu 1360

5392 ANT.....GCGATTAAAGCAGCAATCTGAGCAA 5421

seq\_name: gb\_p14:T24D18

```

seq_documentation_block:
LOCUS      T24D18      80442 bp      DNA      PLN      16-DEC-1999
DEFINITION Arabidopsis thaliana chromosome 1 BAC T24D18 sequence, complete
sequence.
ACCESSION AC010924
VERSION    AC010924.2 GI:6272370
KEYWORDS   HTG
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 80442)
AUTHORS   Liu,S.X., Yu,G., Sakano,H., Jhaveri,A., Lee,J.M., Lenz,C., Pham,P.,
            Toriumi,M., Chin,C., Chlou,J., Choi,E., Chung,M., Gonzalez,A.,
            Howng,B., Koo,T., Li,J., Liu,A., Vaysberg,M., Altafi,H., Brooks,S.,
            Buehler,E., Chao,Q., Conn,L., Conway,A., Hansen,N.,
            Johnson-Hopson,C., Khan,S., Kim,C., Lam,B., Nguyen,M., Palm,C.,
            Shinn,P., Tambunga,G., Davis,R.W., Ecker,J.R., Federspiel,N.A. and
            Theologis,A.
            The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1
            Unpublished
            2 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (28-SEP-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
            3 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (06-NOV-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
            4 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (16-DEC-1999) Plant Gene Expression Center, 800 Buchanan
            St., Albany, CA 94710, USA
            On Nov 6, 1999 this sequence version replaced gi:5931388.
            The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.
            The sequence does not represent the sequence of the entire insert
            of this clone. It is shorter by 1381 bp because we submit only
            the unique sequence of the clone. However, in order to facilitate
            the jointing of overlapping clones in the future for creation of
            larger contigs, we provide small overlaps (200 bp) between
            overlapping submitted clones. The 3' end of this sequence overlaps
            by 200 bp the 5' end of the sequence of the clone F309.
FEATURES             Location/Qualifiers
     1..80442
     /organism="Arabidopsis thaliana"
     /cultivar="Columbia"
     /db_xref="taxon:3702"
     /chromosome="1"
     /clone="T24D18"
     1202..3974
     /gene="T24D18.1"
     join(1202..2102,2184..2365,2479..2568,2682..2969,
     3324..3536,3744..3974)
     /gene="T24D18.1"
     /note="Contains similarity to gb|AF136530 transcripional
     regulator from Zea mays. ESTs gb|F14071, gb|Z26823,
     gb|A1998935 come from this gene."
     /codon_start=1
     /evidence=not_experimental
     /protein_id="AAF18488.1"
     /db_xref="GI:6587797"

```

```

/translation="MSISDEBAEISEIEDYSETPPYRLLRDGTYYKVVNGQLRCPFC
AGKKQDYKELYAHATGVSAGSALOKANHLAMFLENLACVAYEPVPRPV
VPQDDETPNPNHVVYVPMGIYVNPPLKEADKELLDSAYWLQTLKFKPIEVNAF
WVEQDSIVGVIAKFNMGDSWGFAGATELEKEFETGSSKKEWTERGSDSEKAYGCAR
ADDFESQPIGEIYLSKQGLRTVSDISQKNVODRNTVLEELSDMIAMTNDLNRKQYIS
YNTAMSQPIGEIYLSKQGLRTVSDISQKNVODRNTVLEELSDMIAMTNDLNRKQYIS
ESRAKQLEHEALTDLDRQKLEDEKDKDAMNLSQLASRQKKADESVLRKLEVEHQRL
OKDALNKILLKQOLDTKOTLEMEIQELKGKLOVMKHLGDDDDDAVOKKKKMDDEL
DDKKAELGLESMNSVLMTKERQSNDEIOAARKKLIAGLTGLIGAGETDGYKRGSELD
EKPFIDVCKLRISNAVEAATIGCTWQENLKNFSWQPFKHEGFGDGAEEVVDDEDE
OLKLLKREWGKRVHNAVETALVEMNEYNASGRYTPPELWNFKRGRKATLKKEVITFIS
DIKTLKRRKT"
5213..6073
/gene="T24D18.2"
5213..6073
/gene="T24D18.2"
/note="Similar to gi|Q60809 CCR4-associated factor 1
(CAF1) from Mus musculus. EST gb|Z26822 comes from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18489.1"
/db_xref="GI:6587798"
/translation="MSQAPNEEEDDTIEIRVWNHLEQEMALIEQSIDDPYVAMD
TEFGIVCKTVTANPNPNPYSIHYEYNTDLKANVNMKLIQLGLTLDSEKGNLPTCG
TNKOCIMQNFREENVISDMFALDSIELLRKSAIDLEKNNEGVADKFAELLMGSGV
VLYDKIHWTFHCGYDFGLKLLSGKELPEISDFQMEKFFPVVYDIKYLMGFCT
NLNGGLEKIAELLGVKRVGISHQAGSDSLTLRTFIKMKFEFFTSLLKYSGLFGLD
NPRLITGSKN"
6953..7992
/gene="T24D18.3"
join(6953..6960,7349..7649,7771..7836,7933..7992)
/gene="T24D18.3"
/note="Strong similarity to gb|AF067732 ribosomal protein
S12 from Hordeum vulgare. ESTs gb|T41772, gb|T42570,
gb|A1999345, gb|F20068, gb|F20068 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18490.1"
/db_xref="GI:6587799"
/translation="MSGDEAAFPVVPVPPVPAAPEDMDLMTALELTLRKARAYGVV
VRGLHECAKLIKRVQALVLAECNQPDYVYKVKALCADHEVRLTTPSKARTLGWA
GLCKRDSGNARKVVGCSCLVVKDFGEETALSIVNKHIAQ"
8923..13301
/gene="T24D18.4"
join(8923..9036,9346..9549,9665..9846,9939..10065,
10150..10347,10469..10519,10604..11311,11398..11493,
11579..11652,11736..11787,11877..11963,12052..12095,
12184..12258,12341..13301)
/gene="T24D18.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18491.1"
/db_xref="GI:6587800"
/translation="MGPIVATEESKEQALTDAAENLLKPHESTDATLSLLEVMESLL
ATVQDLSVSSVKALHPMRLVADLLRNPDSVRSVSVCLTEIMRITAPENYND
EOMKDIFOVTIEAFKVLADASSYRKAIEVLETVAKVRSSLMIDLECDLLEMFQ
RFLKIRDPHQLVLLVSMETIMTVIDESEVPMDLLEILLTTVKKDSQVSPAALTL
VEKVLSCCTCLQPCIMEALKSSGTLSDMYSPVSSICQSEFATTAQNDVYKPDNEA
DEKISEGOVVPDSLEDLNLGLSRKGRSRKARRGTRRANGDEKVTITANGLSEST
DAFTASGSTRKGMKPKSLMMPPEGYSFTSSKKVQEKELGDSGLVAAKKVPLPS
KVGQTNQSVISSLSSGSRARRSRRTKMEETDHDVSSVATQPAKQTVKKTNPAP
EDLTNRNVKHHEDGIGTKSKKKAADNGLSAKPLAETMMVMPKSGKLIVHSDAK
KNEGASMDPTIPOSSSKKKDSRATTPATKSEQAPKSPHKMRKIAAGEEVESNTNE
LGEELVGRVNVWPLDRKFYGVKSYCRVKMHQVTSYSDGDEALNKKERFKIIE
KDSASEDEDDLLESTPLSAFIOREKSKKRVISKNVSPSSPVRSVMOTKKKDS
VDSIKONKRTKALKAVSNPESTGKNLKLKNGEPDKTRCTGCKOKVTOAMH
RKIKDCDEQDELETKDEDSLKLKESDAEPDRMEDHQELPENINVTETDGEQEA
AKEPTAESKTNGEENPNAETDQKELSKLSEKQETNPVEPKSDGEEQAKPNALKTGYE
NOEAKELTAERKTDDEEHKVADEVEKSKQETNPVEPKSDGEEQAKPNALKTGYE
EESKAEVSTTVLEVD"

```







2194 .....ACTGTGGAGTCTCTAAAGGCACAGCTTT 2221  
 360 isAspProGluGluAlaIleArg.....Hisasp 369  
 2222 CAGACGAGGAGTCAACTAAATATATATCGGAAGAGGTCTCTCCAAAT 2271  
 370 ValIleValSerIleValThraAlaLysLysAspIleLeuValas 386  
 2272 GAGTTGTATCTGAAGAAATTCAAAGCTAAAGATCAACTTCTGTCTGT 2321  
 386 nAspHisLeuLeuAsnPhelValArgGluArgThrLeuAspLysArgTrpA 403  
 2322 GGAAGAAAGTCTGAGGCATCTAGGGAACAC..... 2352  
 403 rgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr 419  
 2353 .....CTGGAGAAAGAA 2364  
 420 AlaLeuGlnSerAlaAlaGlyLysAspAla..... 429  
 2365 AAAACTAAGTTTGGCGCTCTCTGTGCTGACCTTAAAGAAATTTCCCA 2414  
 430 .....AlaLysGlnIleAlaTrpIleLysAspLysLeuLeuH 442  
 2415 CCTTGAAGAGGAGATGAAGAAGTTAAGTGAAGTAGAGATGAGCGCTGC 2464  
 442 isIleTyrTyrGlnAsnSerIleAspArgLeuValGluArgIle 458  
 2465 AT.....AACCTTGATGAAGAAAGGCGCTGGAAGAGAAAT 2502  
 459 PheAlaGlnTyrMetValProHisAsnLeuLeuGluThrThrGluArgMetLy 475  
 2503 GAATCTCAGCTC.....AAGCATTTAGAG...GAGAGTACCAGAA 2540  
 475 sCysLeuTyrTyrLeuTyrIleThrLeu.....AspLeuAsnAlaValL 490  
 2541 AGCAATGAAGCCTGAGGCAAAATTTGGCAGGATCTTGTGCTGCCATTA 2590  
 490 ysAlaLeuAsnGlu.....MetTrp 496  
 2591 AACAGCGAAGAAAGAGAGGAGTGTCTTAAGGTAGTTGATATCTGG 2640  
 497 LysCys.....GlnAsnLeuLeuArgHis.....G1 505  
 2641 AAGSCCAAATATGAGAGTCTCAGCAAAAGATCGCTCAGATTTCTGCCA 2690  
 505 nValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla... 520  
 2691 CATGCAGGAACAAACTGAAGAGCTGAAAAAGACTCATTTCTGATGTTATC 2740  
 521 ..SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536  
 2741 AGCAACTGGAGGGAGAGGAGGAAAGTCTTGATGATAGAGCCAGGCA 2790  
 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLe 553  
 2791 AGTGAACCAAGTCTTCTCAA.....CTGGAGAGATAAATCAGTTAGA 2834  
 553 uGluAsp..... 555  
 2835 GGGTGAATCTGTCTGCAGCCCAAGCGCTGTATCAAAAGAGAGAGAGCTGAAG 2884  
 556 .....AspGluLysIleArg... 560  
 2885 AAAAGAAGCTTGTCTGTCTCTCAGCTCTGCAGAGAAAGAACTCAAGATT 2934  
 561 .....LysGlnLeuGluValLeuValSerProThrCy 571  
 2935 GCTTATCAAGGGGAAAGTAGCGTCTTTCCACCTCGAAGACGCTTAAG 2984  
 571 sSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuG 588  
 2985 CAATGCAAGCAAGACACCTGGATTGT.....CTAGCCAAAGAACTGT 3025

588 lyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeu 604  
 3026 CTGATGAGAAGTATAAGAAAGCAGAAATTTGAAGCCATGGTAAAGTGTTA 3075  
 605 .....LeuGluArgIleAlaProValHis..... 612  
 3076 AAAGAGCAAAATTCGAAAGAATAGCATCCCTCGAATCTGAATGAAAAA 3125  
 613 .....IleAspThrGluSerIleSerA 620  
 3126 CAGTTTACCTGTGTGTCAGAGCGAAAAATGTGAATCTCAAAAGCTTCTG 3175  
 620 laLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspGlu 636  
 3176 GAGAGGTGGAGCATCTGAAAAGGCACTAGATGATTCAGTCAGAGCAT 3225  
 637 AspGluGlyVal..... 640  
 3226 AAGGAAGCAGCTAGCTCAGAAAAATATTGAAATCAACAGCTGATTAATGC 3275  
 641 .....ProThrAspGlnAlaIleArgAla.....GlyLeuG 651  
 3276 AAAAGAAAAAGCCACATCGGACTTGGCTATTAAATCTGAGATGGCGCTC 3325  
 651 luLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAla 667  
 3326 AGCTACAGAAAGCAGTGGT...ACACAAAAAGCGAATTTCAGTGCCTTG 3372  
 668 GluThrPheGluSer.....LeuLeuAla..... 675  
 3373 CAGAAATGAGCTGTCTCGTTCTCTGTGATCTCTCGCTCTAAAGAGGGGA 3422  
 675 ..... 675  
 3423 AGTAGAAGAGCTTAATAAGGAGGTGCCCTAAGACAAGAGAGAAATACAGC 3472  
 676 .....CysLeuLysMetAspAspGluLysValAlaGluAla 687  
 3473 AGCAGCAACAAACCATACTAAGCTCAGGGAAGAAACGGCATTTGGCT 3522  
 688 AlaLeu.....GlnIle 691  
 3523 GCACCTTAAGATAAAGTCGCCCTGCAGGAAAAAGAGATCAAGCAGCAAT 3572  
 691 ePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgS 708  
 3573 CCAGGCCACTAAAGGTGCT.....GAAAAGAGATGGCTAAATTTGAAT 3616  
 708 erAlaLeuLeuProValLeuHisLysSerLysLys..... 720  
 3617 CT.....GTAAATTTCTGAAAAATCAAAAAGAAATTTGAATGTTG 3654  
 721 ..GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI 736  
 3655 GAACAGGACATCCAGAACCCAGAGAGAGACCTTTCTTCATTCAGAGCA 3704  
 736 ePheSerSerLys.....GluThrGlnPheAlaGln..... 746  
 3705 GCATCAAGCAAACTTGGTGAGTCCCAAGGCTTCAAGCTCTTATTGCAG 3754  
 747 .....IlePheGluProLeu 751  
 3755 ATCTGGAAGAAAGTGCAGAGCAGAGAACTGATTTGTGAGGCCCAA 3804  
 752 HisLys.....SerLeuAs 756  
 3805 AATAAGGCTCTGAGGCAAAACCGTTGGCCTCAGAAAAAGGCTTCAGTCTC 3854  
 756 pProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773  
 3855 AGAGAGCAGTTAGAAAGAAATCCAAAGCTTTAGAAATTTGAATTCGTAAG 3904















```

52 GluLysGluLeuTyrLeuAsnLeuAlaLeuHis.....LeuAlaSerAs 66
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4734 CAATCAGATTACATTTTACGATATGCATCAGAGTATATATATGTTATGA 4783
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
66 pPhePheLeuLysHis.....ProGlyLysAspVala 77
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4784 TCTTTTATAAACACATGAAATAAATAATCTAGCGATGGAAAAAATTTT. 4832
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
77 rgLeuLeuValAlaCysCysLeuAlaAlaPheArgIle..... 90
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4833 .....GATATTATAAAGGTTAAAGAATG 4856
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
91 .....TyrAlaProGl 94
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4857 ATAAATAAACCAACCAACAAAGTTTCAATGAAGCTATGCAATATGCTAATAA 4906
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
94 uAlaProTyrThrSerProAspLysLeuLys.....AspIlePheM 108
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4907 AATGGATGAAAAAATAAGGACCATTTATAAANAATAGAAAAATCAACTTTATA 4956
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
108 etPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGln 124
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
4957 ATTTAATAATAATAACATCGTTCATTAAGGGGTAATAATATGAAAAA 5006
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
125 PheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerty 141
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5007 GTTAGGAAA.....CAAGCAAG 5023
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
141 rAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeut 158
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5024 AAAT.....GCAATTGATGATATAAATAATATACATTTCTAATATT. 5063
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
158 yrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHis 174
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5064 ..AAACGATTTAACC.....AAATCTAAAGAACGA 5093
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
175 MethHisMetValAspLeuMetSerSerIleCysGluGlyAspThrVa 191
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5094 TTAGATCAGATTAAAGAAACCACTTAACATTAAGAGAGAGTGATGTTT 5143
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisL 208
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5144 AAATAAT.....GATAAAACCAAAATAGCTTATATACATACAAA 5184
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
208 ysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu...LeuLys 223
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5185 TAAATACGGAAGATAGATCTAATTTTATAATATATATAATATGAAA 5234
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
224 ArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVa 240
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5235 CATAAATAGATAGTACTATCTTGAATAAAGCTATGCGATTATATGAATGATG 5284
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
240 lLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheA 257
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5285 ATCAAAATCTGCACAGATGTTTATTAATATAGATCTTGAATATGAACG 5334
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
257 spLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSerVal 273
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5335 AT.....ATATATAAATAGATAGAAAGATCTTTTAAATAATATT 5372
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
274 LeuPro.....GlnLeuGluPheLysLeuLysSerAsnAs 285
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5373 TTAAGAGAAAAACAGAAATATGGAGGCAGAAATATAA...AAAATGAATGA 5419
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
285 p.....AsnGluGluArgLeuGlnValValLys..... 294
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5420 AATGTATAATACGTTAATGAACAGAAAGAAATAATAAACAATATAAA 5469
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
295 .....LeuLeuAlaLysMetPheGly 301
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5470 AAAATTATGAATAAGAAATTATGGAACATATAAAAAAGAAACAAATGAA 5519
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
302 AlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrp..... 315

```

```

   ||| :::: :::: :::: :::: :::: |||
5520 AAAAAAATAAATTTATGGAATCTAATAACAACATCAATTAACACTTTAAT 5569
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
316 .....GlnCysTyrLeuGlyArgPheAsnAspIleHisV 327
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5570 GGATTCATTCAGATCTATGTTTATAATGAATATATAAATGATTATAATT 5619
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
327 alProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsn 343
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5620 TAAATGAAATTTTGAA.....AAA 5639
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
344 HisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerH1 360
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5640 CATCAAAATATATTGAATGAATATATAATGATTAAATGAA..... 5681
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
360 sasProGluGluAlaIleArgHisAspValIleValSerIleValThrA 377
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5682 .....TCATATAATATTATTAAACAAAAATGACTG 5712
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
377 laAlaLysLysAspIleLeuLeuValAsnAspHisLeu...LeuAsnPhe 392
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5713 AA.....ATTATAATGATAATTTAGATTATAATGAA 5744
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
393 ValArgGluArgThrLeuAspLysArgTirPargValArgLysGluAlaMe 409
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5745 ATAAAGAA.....ATTAAAGAAGTAGCACA 5770
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
409 tMetGlyLeuAlaGlnIleTyrLysLysTyr.....AlaL 421
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5771 AACAGAAATGATCAAACTTAATAAAAAAGTTGATGAATTAATAAATATT 5820
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
421 euGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTirPileLys 437
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5821 TGAATAATATTAAAGAACAGAGGACATCGATTAAATGATTATATAA 5870
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
438 AspLysLeuLeuHisIleTyr.....TyrGlnAsnSerI1 449
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5871 GAAAAAATATTTAACCTTATATAATAAATGTTCAACAACAATAATAAT 5920
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
449 eAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProH 466
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5921 AGATGATCTTATAATTTATATACAGTTAAAAAACAGTAT..... 5960
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
466 isAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeuTyrAla 482
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5961 .....ATTAAACTATTGAAGATGCGAAATTTTA..... 5990
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
483 ThrLeuAsp...LeuAsnAlaVal.....LysAlaLeuAsnG1 494
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
5991 ...TTAGATTCATTGAACACAATAAGAGAAAAATAAATCAGTAGCAAA 6037
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
494 uMetTrpLysCys.....GlnAsnLeuLeuArgH1SG 505
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
6038 TCTAGAAATTTGTACTAATAAAGAGATATAAAAAATTTACTTAACATG 6087
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
505 lnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSer 521
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
6088 TTATAAG..... 6095
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
522 ValLysAlaIlePheSerLysValMetValIle.....ThrAr 534
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
6096 ...TTGCAAAATTTTTCAGGATTATTGTTAATGCTGATACAAATACGGA 6142
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
534 gAsnLeuProAsp...ProGlyLysAlaGlnAspPheMetLys..... 547
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
6143 AATAACTCCAGAAAATCCTTTAGAGAGATAATGATTATTATTAATTTACAAT 6192
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
548 .....LysPheThrGlnValLeuGluAspAsp... 556
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
6193 TATATTTTGAAGAAACATGAAATACATCAACATTGGAAAAATGATTCT 6242
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||
557 .....GluLy 558
   :::::::::::::::::::: ||| :::: :::: :::: :::: :::: ||

```







14043 CTTCAAGAAATTAAGACTATTCAACTTCGGATGGAAGTTCTCAAGAGAT 14092  
 722 OProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerS 739  
 14093 TTCATGAATCAAGTCAACTGCTGAAGAAATTTACTTCAAGAGCGGTTCAAA 14142  
 739 erLysGluThrGln.....PheAlaGlnIlePheGluProLeuHis 752  
 14143 ATAGTGAACAACAATGTAACATTTCAAGAAATTTGCAAAATGTTCAAT 14192  
 753 LysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeu..... 767  
 14193 TGGGAA...GAACAAGCTTTACAGAGAAATCATCAAAATTTGAGAAT 14239  
 768 .....ValThrIleGlyHisIleAlaL 775  
 14240 TATATGACTGCGATTCAGATTCGAAGTTCGAACACAGTAGGA.....AGCT 14283  
 775 euLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThr 791  
 14284 TGAAGGCTCCGATGATCAACAGCTCTCCAGATTCATACATTTTCATCA 14333  
 792 PheIleValLysAspLeuMetAsn.....AspAr 802  
 14334 ACAAGACGACGAGAAAGCTGTATATAAATTTGATGCGACGAGCTGATAACGC 14383  
 802 gLeuProGlyLysThrThrLysLeuTrpValProAspGluGluValS 819  
 14384 TATTTCACATCATCCACCTTATCAATGATGGAATCGGATCGAATGCTTC 14433  
 819 erProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835  
 14434 TTCATAAACTATTCTGAACACGATGCTCTT ..... 14465  
 836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuAr 852  
 14466 ...TGTAATCTGACAACTCTGCTTCAAGATCTGAACAGAAACAATGAC 14512  
 852 g.....LeuLeuThrThrIleLeuHisSerAspGlyAspLeuT 865  
 14513 TCAAGGATTTCTACAGAAAGTGGAGAGTGTCAACAGACAAGAAATAG 14562  
 865 hrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 881  
 14563 CTGACGACCTTAATTAAT...TGTAATGAAGTACTTGAATTTCT 14609  
 882 AlaGlySerAla.....IleValLys..LeuAlaGlnProCyst 895  
 14610 GATGACGAAGTTCATGGTATCTGTTCAAGTCGTAAAGAGACAAGATCAT 14659  
 895 yrHisGluIleIle.....ThrLeuGluGlnTyrGlnLeuCysAla 908  
 14660 CACAAAAGATTTCTTTGGCAAGTGAACCTTGAACAAACCAACACTCAACA 14709  
 909 LeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLy 925  
 14710 TTGGCAAGTGGCATGAA.....ACGAATTC 14735  
 925 sLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaI 942  
 14736 CATTTATACGGAATTCATCATTCGATCGGTACCCGAGAAGAAATTTGAGAAT 14785  
 942 leCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAla 958  
 14786 TGGTTTCTATCGAGCTTCGGTATGATTGTAGAAAAAAGTTTGGAGTATCA 14835  
 959 ArgGlnCysLeuValLysAsnIleAsnVal.....ArgArgGluTyr.. 972  
 14836 GAGGAAGTTTAACAACACTCTTAATTTGATTTCCAAAAATGATGGAG 14885  
 973 .....LeuLysGlnHisAlaAlaValSerGluLysLeuL 984  
 14886 CAGTTTGACGATGCCAATTTGCAAAAAAGAAATTTGGTTATCAAAAAATCTGG 14935

984 euSerLeuLeuPro...GluTyrValValProTyrThrIleHisLeuLeu 999  
 14936 CATTTGCTTTTACCTGCGAGGCTTAATCATGAATCGGTGGCAGCAT 14985  
 1000 AlaHisAspPro.....AspTyrValLysVa 1008  
 14986 CAGGCACCAACAGAGAGGAGCATTGCGAGAGTTGGATTGAAACAAGC 15035  
 1008 lGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValL 1025  
 15036 AAGAGAAGCAAAAGTTATGATGGATATTCAGAAATGC.....GTAT 15076  
 1025 euGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArg 1041  
 15077 TCACACATGTCAGGATACACTGAAATGAATAA..... 15111  
 1042 LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAs 1058  
 15112 .....CCAGATGA 15119  
 1058 pAlaLysMetAsnGlu.....LysLeuTyrThrValCysA 1070  
 15120 TGCTGAAATTTACCACATCTCTGCTCTTCTTGTGTACTCTGTCAATCTGTG 15169  
 1070 spValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlu 1086  
 15170 ATCTC.....GTGCAATGGCTCTTCAAA..... 15195  
 1087 SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLy 1103  
 15196 .....ATTCAAGTTGACAG 15209  
 1103 sAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPheP 1120  
 15210 TAAATATGACTACTATAGAGCGCAGCTCCAAAAAATGCAGAGACTACAT 15259  
 1120 heThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136  
 15260 TCACGAATCCAAT.....GCAGATGCATTTTCAATG 15291  
 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetG 1153  
 15292 GCTCTTCAAGAGCTGTGAG.....GTTACATCTTCTGTTATTTGGAG 15335  
 1153 uThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgI 1170  
 15336 TAGCGTTTCAAACTCAGATGCTCAAAACTACAGTTTCTGCACAAGTTGA 15385  
 1170 leLysGlyArgLeuAspSerSerGluMet.....Asp 1180  
 15386 TTCT.....ATTCTAAACCGAAATGCTATGAAGCGCAGTCAGTGAA 15429  
 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLy 1197  
 15430 AAGCTGTAATCAGGATCACTCTTG..... 15456  
 1197 sSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlyLysP 1214  
 15457 .....AAGAAAGATTCCATCGAAGCAGTA.....GAGATGAACATTC 15493  
 1214 roArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGly 1230  
 15494 CGATTTCGCAAAAGAAAT.....CTTCAGAAAACTATTCA 15531  
 1231 MetAspAspLeuThrLysLeuValGlnGluLysProLysGlySerG 1247  
 15532 ATTGATCGATCAAACTCTACTGTTGAATGCAA..... 15564  
 1247 nArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnT 1264  
 15565 .....GGACAAATTCATCGGAAGACATTTGCAATTCAT 15598



```

215 yrAspLeuAlaLysAlaLeu.....Leu 222
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12376 TAAATCTCAGAGAAAGTTTGTAGATCAAAACGGTCAAAATGACGTCTTCGCTT 12425
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
223 LysArgThrAlaGlnAlaIleGluProTyrIleThrThr..... 235
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12426 ACCGATCAGATGAGAGATCGAAGTTTAAAGTTTCCATATATATATCTCTTT 12475
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 .....PhePheAsnGlnValLeuMetLeuG 244
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12476 GATTTCTTCGAAAACTTTGGAGATCTGATATCACAAGCGTTCAACTTG 12525
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
244 LysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12526 CTCAACCAATATGCTGATGAGACAGAAAAATCATTTGAA..... 12566
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
261 GluLeuTyrAsnIleAspSerHisLeuLeu.LeuSerValLeuProGlnL 277
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12567 .....ATTCCAAGAACATCTCTTAAACAGAGTTGTGCCAGA 12604
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
277 euGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12605 AACATCGGAAGAATCGATTCATTCATCTCGAAGACATCTCAACTTGCTG 12654
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGI 310
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12655 AAGAAGCTTCGTCACATATTT...CTGAAAGCTATCATTTGCAATCA 12701
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
310 nAsnLys...ProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleH 326
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12702 GAATTTGTAGTCTCTGCTCCAAAACAAGTTA.....GTAC 12736
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
326 isValPro.IleArgLeuGluCysValLysPheAlaSer.HisCysLeuM 342
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12737 ATCTCTCACATTTGATTACCGGGAAGATTTTCAATCAAAATTCGGA 12786
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
342 etAsnHisProAspLeuAlaLysAspLeu..... 351
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12787 TTATATTCGGTGTATCTGACAAGAGATGCTCAAGCGCGCTTTCTCAGTA 12836
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
352 .....ThrGluTyrLeuLysValArgSerHisAspPr 362
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12837 TCAGACGAACATTAATACATTTATCTCGCAATTCGAGCAATCGATG 12886
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
362 oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL 379
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12887 GAGTCAGATCAAGTTATCTTCCAAGCAAAATCTATATATCTGCTAACA 12936
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
379 ysLysAspIleLeuLeuValAsn...AspHisLeuLeuAsnPheVal... 393
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12937 TAAATCTCTCGCAACATCTAATCTAGACACCCCTCATTTGGACAACGTATG 12986
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
394 .....ArgGluArgThrLeuAspLysArgTr 402
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12987 CCAAGAGAACGAGATTCGGAGCAGAGTGAAGCTTCATTCTCTCAACGAAG 13036
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
402 pArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle..... 415
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13037 CAGTGTGAGAAATGGAGAACACTTTCAAAATTCAAACAATATATTTTGATA 13086
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
416 .....TyrLysLysTyr..... 419
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13087 CAGTTGTAGAAAGTTTGGATCCGATCGAAGAAAGGCATCAATAATAAT 13136
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
420 .....AlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAl 434
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13137 CTCATTGGTCTGCTGACTCTGTATTTGAAGGCAGCTATCCAAGTTGC 13186
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
434 a.TripLysAspLysLeuLeuHisIleTyr..... 444
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13187 TTCTTTGAAGACTACCTGATTCACATCTCGATTCACAGCTTCAATCACTC 13236
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 .....TyrGlnAsnSerIleAspAspArgLeuLeuValGluAr 457
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

13237 CAGTCTTCAAGATTCAAGAACATGTCGTCAGCCGACTG.....AACAG 13280
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 gIlePheAlaGln.....TyrMetValProHisA 467
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13281 GTTGTTCAGGATTCGTGTCAAAGGAAGACATCCAAATGAACCGTGA 13330
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
467 snLeuGluThrThrGluArgMetLysCys...LeuTyrTyrLeuTyrAla 482
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13331 AGTTG...TGAATAGAAAAATTGAGGTGCTGAAGAGTATTTTGAATTGTT 13377
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
483 ThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys... 498
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13378 ATGCTGTGCTGAATGCCATAAGAGTTTGGAGAGAAGCCTGGAGAAGCCT 13427
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
499 .....GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspL 512
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13428 TCTCAGCAGAGTGAACACTTGTGCGACCCGAG.....GAGTTACACAAGT 13471
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
512 euIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13472 TGTTCAGAATGCGTTT.....GCTGTAACGGAGACAATCGTATTCCAAG 13514
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
529 ValMetValIle.....ThrArgAsnLeuProAspProGlyLysAl 542
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13515 ATTCGTAAGGTTTAGGATCAATGCAAAACTTTGAAATACCAAGAACAGT 13564
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
542 aGlnAspPheMetLysLysPheThrGlnValLeuGlu.....A 555
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13565 AAAAGAAACCATTTCAAAAAATTTGAAATTCGAAATTTGCCACCAGTCG 13614
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
555 spAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCys 571
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13615 ACACATTATTGTTATCAAAATATAAAGATCACATGTCAGCAAGACAGTC 13664
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
572 SerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGI 588
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13665 GAA.....GAACAAGAATTGTTTCAGTTCAATGTT 13693
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
588 yAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuL 605
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13694 TAATTTCAATCGGCACACTCAATTCGAAATTCGAAATGTCACAAAATGTTT 13743
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
605 euGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeu 621
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13744 CAAAATGCAATTCGAAGAGAAACAGTCGTTCTTCCCAACCTTGTAGTATT 13793
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
622 IleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGI 638
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13794 TTAGCTTCTCGGAATCTTGTACAATT..... 13820
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
638 uGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysv 655
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13821 .....GTTCTCTAGAGAGCAGACAACCTTTCCGATTTCAGAAAATATTCTCA 13866
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
655 aLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGlu 671
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13867 ATTTCTCGATACAACAGGAAGATTCGAGCGATCAGTCGCCACAGCCGT 13916
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
672 SerLeuLeuAlaCysLeuLysMetAspGluLysValAlaGluAlaAl 688
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13917 .....GATGAATAAACAATTTCTAAGGCATTA 13942
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
688 aLeuGlnIlePhe.....LysAsnThrGlySer..... 697
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13943 TCTGCAAAACATTTGCCATTCAAAAGGCATCTTCAGAAAGTCAGCAGATAA 13992
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
698 .....LysIleGluGluAspPheProHisIleArgSerAlaLeu 710
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13993 TAGAAATCATTAACACTTCTCAAGTTTCCATTTGGTTGGTCACACGACAATG 14042
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
711 LeuPro.....ValLeuHisHisLysSerLysLysGlyPr 722
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



GKLNPEPQSYVDDQIISHTQLEVIANIKSAVSVISNDVLSKRSEDEKAVISNIA  
 NLIISCDLASSTNSFEFQDILEPQANVLLGTPASQVETRNLEPQIOSEQDFW  
 STNSQEKASLVIRKLVKYDAMKMAFEIAYQSLNSNITSEETNLEPSKQSFDIR  
 ESIMASGVSSSEHVQKTEILSKSELSTIVLPAVEYEAQVNIKALSEPNCESILG  
 KUNADPQSAVMDIIMNEQKUTVSNQISQASIESVISNDSSLEKLETPEDVKSILKISE  
 IIVSHDLTALSQSCDFNLLQSTNQEKADIEFKTPNSOMLIEREPTEKROQVQFWS  
 TASNEKQEMFKERVETIHAMLOTFASLSETVORDFMATVQSLAAQLQLTPRE  
 ILCAAFGISEQIEOTFGLDEVDWSDHIDPCALRDNLNVLNRIVNADAPNVGNLIS  
 REDNEQETVLEEKNRIQFLLNORSLODQVMTSSLTRSDERIEVKVSNIIISLSS  
 ENGLDISQAVLQPNMSDETEKSEFIPRNILLNRVVPETSEESIQFHWKZSLAAE  
 ASSTISEKLSMQSEFVSAAKOVSTSLUDIRKIFNQNSEIIFGDTRVQVKAAS  
 VSEDTLNTLSSQESDQSIQKLSQKQSLKSLATSLNLDLIGOLMPREKADSE  
 QSEASLRSFLEMENFKIQIIFEDTVVRKFGSDEERASIIINLGLLTSVNLKAAIQ  
 VASFLRPDLSOSTASITVPFKIQHROPTQVQGVQSWKERHPNETVEVYVNRKIEV  
 LKASILNCVAVASEKLESLEKPSQOQDETCATRGTEVQVNAFAVTEQOSKILKVL  
 GSNQFPEIPEAVKETISKNLKLINLPDVTLLVSKYKESHVSKTYEEQELVQPMNFK  
 SALNSEIEMQNSQKIEETVVLNPLVAILASANLATIVEPSESQSLSDSENILNFSI  
 QHEESERASVNDREINILRHLYQTFQAKASSEYSRVIEIITKTSQSHLVHTMLEE  
 IKTIOLRMEVLKRVSMNOSTABELLQDEAFQNSEQOMETFORIVQVNHWEETLIEII  
 RNLINMTAIPFESTVSGLKAPDQHEASDSYISSTRAEAVINLMAADNAISTSS  
 VYRSSQVRAPESEESTOKTEALRRESLESNAKTFVDRRRENVSLSRKASVERETNM  
 EARVAKAQSIPEVTLQSKKQOESIFSSVVEKSDLNVCQSMTPAKPIGAKVSLQTK  
 VKEVETISATNVASAVECEVGLSKREQSRDGTSMVRKASIEBEREFGEVETISTE  
 KYLEKRDQLESNIKSIPEEVEESAEGFNDEVIIGGVMSGLEAPLEQEEETKLLTM  
 KRSASERASLKAETPEEVEADEFEKIEADQELTAVQELVKATGILNAAASRETNAS  
 SMKESYSPEDVIELSTESRRQNSQSPKQKEMEEIVGLWNTGTKEGNASKVLPKH  
 PIDTASMKAAKAAKONSIMTGSLOKSPSAEAMGIVSOKVTEGANSKFGIACQAVETT  
 LTASQSGVTCRDISVSNIGVASVETVTOFNSKETGIGPCASNLIAPPKASMETETGK  
 ISNLSQTSRLNKMAASDVGTTVESKTOAQGDNDVLSLQKVASSDAIKAMQASRDSA  
 ISVDFQRDRETSKEDLNFKSTNSQTKLFLSKESKEESGIFIRSHXEYBETOKTL  
 RHRASRESASTVAPTNOEQVMNFDDKVEDSVAEGSLSIGIVRESSQSEYVQHAER  
 TSELRKLSMNEEAVAGVAVSETNETFRGYOQGDVEVOTGAAMGRIEAPRPOGRGAEI  
 TOKLARTLSVERSAKASMAESQVTOIQKREDSLASEYSVRDTLLKSSVSHVATE  
 OMSEHLVMSKSEHHSLEKLOERLIEKESFSQEFISENOGVHTHWDVIDNNGDALIC  
 WKSSETEQSLDAKQVTERLACTLDSLVRLLGGTDEEKAIEHVIVGASEVLNISETR  
 ADMMSRQDASDDATVYASDILERSLSTVHEFGESEASTFFGVGLVYKPAKEEV  
 GHSFSETRKLSFDITAISETTDMDSIELRPECDSQVTLISHKNOQKQKDLKAT  
 VENTADQVTLLEKREVEQSDVKKKEFEDEIEVLPIMRWAETLSMKASTSVEIQL  
 ANKLKSPQAQESOKRYKLTANLEINKINVNAHEVMSATTALCEKTSQGVDSATRL  
 ROKSERVEKQYQENQWLLSTNAEWTLLNDEDSVTIAQSVQDSMAFVSASATVN  
 LNSDTSLIKAQPEIGLOKSIQSQNEKTVQGFSSATIGNELLYLRLGHLDLELEFLVD  
 EINRQAVGGRIRERFGKSTGGTGIILVRALPKVRETSTHVTIVATSFQIIFSTMSAG  
 DEISEANVELTIPTSVGAIEKSTVARSDSMTFTSHASEYTAFTADVADIVASAS  
 TAARKKAIPTERTSOKLEVGADGIEILSLWEGIETDLDTQQLDVLRLKRSLSQTIIE  
 SSEEVERINOYLEMPEEKGLVHLINQVONKTCERNFVASICSLDVTSHRSKIEIPPTY  
 SVSKVLTERKVRLETRWRVTESGDVRFNVAVINLHRYLSLKPTLAQETVLREVVRIAAOP  
 LFTSAGDIESDVVNSHQLEKTFVREEDRTIIGARKGENIKKKEEAGEKVKLVE  
 LGGLNLPNEICDLSWKIPRLGLGVKLETEFEFEDEFLWNSQLQADRF\*

gene

21201..27586  
 /gene="B0228.2"  
 /join(21201..27586,25621..25699,25774..25874,25950..26055,  
 26445..26576,26630..26780,26826..26900,26948..27083,  
 27129..27228,27467..27586)  
 /gene="B0228.2"

/note="coded for by C. elegans cDNA yk19cl.5; coded for by  
 C. elegans cDNA yk45a10.5; coded for by C. elegans cDNA  
 yk21e5.5; coded for by C. elegans cDNA yk21e5.3; coded for  
 by C. elegans cDNA yk35cl.3; coded for by C. elegans cDNA  
 yk45a10.3"

/codon\_start=1

/protein\_id="AAC38806.1"

/db\_xref="GI:726363"

/translation="MRTKESSEDIVGVGIAYNVNPEENLISAIKLDEKASGGCYELTTK  
 AAGDEVKTIQMSILSKTQAVEKVKGMIOKVAKDEIRVLETTESVNAFNFEIPEEN  
 FKIGTAKTCANNAAPTMLHECLDFYKMFYMNKKDQDFSLIKITILINSIESKTL  
 SCTAAEHIEATRNPFNPAEFSIQKTVVDYINKIEGASLNTFEALNEXKAMTMSLR

alignment\_scores:

Quality: 234.00 Length: 1593  
 Ratio: 0.312 Gaps: 78  
 Percent Similarity: 47.081 Percent Identity: 19.397  
 alignment\_block:  
 US-09-512-581-2 x CELB0228 ..  
 Align seg 1/1 to: CELB0228 from: 1 to: 38946  
 11 GlyLysIleThrTyProGlyValLysGluLeuSerAspLysIleSe 27  
 |||||:||||: ||| :||| ||| |||  
 11613 GGAAACCTTAATGACACAGATCCCAATACAGTAATGGACATAATAT 11662  
 27 rLysGluGlu.....MetValArgLeuLysMetValValL 40  
 :|||:||||: ||| :||| ||| |||  
 11663 GAACGACAAAGAAATAACAGTTGCTCCACATACAGTCAGCCATTG 11712  
 40 yThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56  
 :|||:||||: ||| :||| ||| |||  
 11713 AATCGGTAATTTCCGAATGACTCATCTCTTGAACAACTACGGAAGTTGAT 11762  
 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGl 73  
 :|||:||||: ||| :||| ||| |||  
 11763 AAATCTGTGATTCTTAAATA...TCTGAATATTACGTTTCTCAT.... 11804  
 73 yLysAspValArgLeuLeuValAlaCysCys..... 83  
 :|||:||||: ||| :||| ||| |||  
 11805 ...GATTTGACTGCATGACGACGATGCTGTTCAAGTTTCAATTCGCTTC 11850  
 84 .....LeuAlaAspIlePheArgIleTyAlaPro 93  
 :|||:||||: ||| :||| ||| |||  
 11851 AAGAAATCAATACTCAAGAAAGAACGACAGATATTC...TTAAACATCCA 11897  
 94 GluAlaProTyThrSerProAspLysLeuLysAspIlePheMetPheIl 110  
 :|||:||||: ||| :||| ||| |||  
 11898 AATTCTCAAAATGCTCATT...GAACGCTCGCGAGCCA.....AT 11935  
 110 eThrArgGlnLeuLysGluLeuGluAspThrLysSerProGlnPheAsnA 127  
 :|||:||||: ||| :||| ||| |||  
 11936 CGAAACCAAGTACAGGATTTCTGGAGTACTGCAAGCAATTTCTGAAAGC 11985  
 127 rgTyThrLeuLeuGluAsnIle.....AlaTrpValLys 139  
 :|||:||||: ||| :||| ||| |||  
 11986 AAGAAATGTTCTTCAAGAAAGAAAGTTGAAACAATTCATGCAATGCTTCAA 12035  
 140 SerTyThrAsn.IleCysPheGluLeuGluAspSerAsnGluIlePheThr 156  
 :|||:||||: ||| :||| ||| |||  
 12036 ACATTCAGCTTCTCTGGTGTAGTCAAAACAGTTCAAAGAGATTTATGCGC 12085  
 156 InLeuTyArgThrLeuPheSerValIle...AsnAsnGlyHisAsnGln 171  
 :|||:||||: ||| :||| ||| |||  
 12086 AACTGTACAGAGCCTTCTGCTGCTTCTGTTCAATCCAATTGACACTCGAG 12135  
 172 LysValHisMetHisMetValAspLeu..... 180  
 :|||:||||: ||| :||| ||| |||  
 12136 AAATTTTATGTGCAGCATTTTGGAAATATCAAAATGAACAAATCGAACAAACA 12185  
 181 .....MetSerSerIleCysGluGlyAspThrV 191  
 :|||:||||: ||| :||| ||| |||  
 12186 TTCGCGGCTCTCGATGAAGTTGACTGGAGTCATATA.....GATA 12225  
 191 alSerGlnGluLeuLeuAspThrValLeuValAsnLeu..... 203  
 :|||:||||: ||| :||| ||| |||  
 12226 TTCGCTGTGCACACTCAGAGATAATTTGTTAGTCAATCTGAGAAATGTAAT 12275  
 203 ..... 203  
 12276 GCTGATGCTCCAAATGCTTTTGGAAATTTGATCATGCTGCTCCAGATGAAA 12325  
 204 .....ValProAlaHisLysAsnLeuAsnLysGlnAlaT 215  
 :|||:||||: ||| :||| ||| |||  
 12326 CCAAGAAACGCTTACGGTCTTCTTGAAGAAAAAATTCGCATTCAATTCATTAC 12375



```

13769 AGGAGAGATAAAATCAATCGAACAACTCGGAGAAATAGACGTGGTCTT 13818
1336 .....AsnThrGluGlnLysSerLysSerLysGlnHi 1346
      |||:|||||:|||||:|||||:|||||:|||||:
13819 CGAAACAGCTAGAGCTAGGTTCAAGAGAGAAACAAAGGAATGGGAAA 13868
1346 sArgValSerArgAlaGlnGlnArgAlaGluSer...ProGluSerS 1362
      |||:|||||:|||||:|||||:|||||:|||||:
13869 CGAGTATACAAACAAAATATACAGAGAGAACTCAAGATGTCACCACTCAA 13918
1362 eAlaIleGluSerThrGln 1368
      |||:|||||:|||||:|||||:|||||:
13919 GTCATATCTCAGATGAGAA 13938

seq_name: gb_in3:CELB0228

seq documentation block:
LOCUS      CELB0228      38946 bp      DNA      INV      30-MAR-1995
DEFINITION Caenorhabditis elegans cosmid B0228.
ACCESSION  U23168
VERSION    U23168.1
KEYWORDS   U23168.1 GI:726361
SOURCE     .
ORGANISM   Caenorhabditis elegans strain-Bristol N2.
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 38946)
            Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Showkneen,R., Smalton,N., Smith,A., Sonhammer,E.,
            Staden,K., Sulston,R., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
            2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
            Nature 368 (6466), 32-38 (1994)
            94150718
            2 (bases 1 to 38946)
            Leimbach,D.
            The sequence of C. elegans cosmid B0228
            Unpublished (1994)
            3 (bases 1 to 38946)
            Waterston,R.
            Direct Submission
            Submitted (21-MAR-1995)
            Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RQ, England
            e-mail: rwenematode.wustl.edu and jesesanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is B0495, 8001 bp overlap; 3' cosmid is C06A8, 750 bp
overlap. Actual start of this cosmid is at base position 4495 of
CELB0228; actual end is at 6105 of CELC006A8

NOTES:
Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in
preparation).
Location/Qualifiers

```



```

795 ysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12323 GTGATCACTGGAGACTCTCG.....AAGAAACGAAGAAGT 12363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
812 TrpValProAspGluGluValSerProGlu..... 821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12364 TGGTCCCTCCAGAGAAGTCTTTATTAGAACAGCTAGATTATCGAATTC 12413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
822 .....ThrMetValLysIleGlnAlaIleLysMetM 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12414 TCGATTGAGGATTATCTCCCAAAATAACTATTGTATGATCAAAATAC 12463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 etValArgTrpLeuLeuGlyMetLysAsnHisSerLysSerGlyThr 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12464 AAATCTACAGCTGCGGCAAGAGTCAATAATTCGACAAACGGACCT 12513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 SerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuTh 865
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12514 GGTGTGAATAATATTTAATTACACTACGTGCGGAAAGGGATATCTTGA 12563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
865 rGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12564 TACAAAGTGCAGTGGCTGGAAGAGATGCAAAATCTTGAGACAAAAAA 12613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
882 laGlySerAlaIleValLysLeu.....AlaGlnGluProCysTyrHis 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12614 TTTCTTTGATGATGTGCAATTACAAAGTCTCGTACTAGTACATAT 12663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 GluIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAs 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12664 TCAAGAGTTGAAAGGAAATCAATCTTCCATTTTCAACAGCATGACGA 12713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 pGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyL 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12714 C.....ATTATGAGAAAT 12727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
930 euSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12728 TAATCAATTAATCTA..... 12744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
947 AlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVa 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12745 .....TTAAGAGAAAGTAACATACATTGCGGAATGAGCTGA 12782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
963 lLysAsnIleAsnValArgArgGluTyr.....LeuL 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12783 AAACAACAATAACAAGAAAGAGGAACTGCAATCTGAATTAGATAAATTGA 12832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 ysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyr 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12833 AGCAAAATGTTGCGCTATCGAGTCCGAATTCAGACGCTTG..... 12873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
991 ValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLy 1007
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12874 .....AAATATTCTAT 12884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1007 svalGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheV 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12885 GCAAGAAAGAGCAAGAGCTCAAAATTAGCTTAAAGAA..... 12921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 alLeuGluIle.....LeuMetAlaLys 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12922 .....GAGTTCATCTGGTGGAAAGCGCTCACAGACATATTGGAGAA 12966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMet.....ValG 1045
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12967 CATGAACAATTTGAGCTCAAGCGATTATGAGAAGCTAGAAAGCGAGATAGA 13016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1045 uAsnIleLys.....GlnThrLysAspAlaGlnGlyProAsp...A 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13017 AAATTTGAAGGAGGAAGTATAGAAAAATAAGGCGTCAAGGCGGAAGCCG 13066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 spAlaLysMetAsn.....GluLysLeuTyrThr 1067
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13067 AGGAAAAATTTAACAGGCTGAGAAGACAAGCGCAAGAGATTAAAAACA 13116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1068 ValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSe 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13117 .....TCAAAACTCTCACAGGACTCATTTGACTGAACA 13148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1084 rLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnP 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13149 AGTAATAGTCTAAGGATGCA..... 13170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 roAspLysAsn...PheSerAsnThrLysAsnTyrLeuProProGluMet 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13171 .....AGAACGTGTTGGAATAATCTTGTAGTGAGGCAAAAGCGGAGATC 13215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAl 1133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13216 GAAGAGTTACAAATGCAAAAGTAGCACAAAGTAACCAACCAAGTTAGAAGC 13265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1133 aValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13266 AATAAGAAAAATTACAGAAGACGCAAAAGGTTCAAGAGAGCTTCAAG 13315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 erArgMetGluThrValSerAsnAlaSerSerSerSerSerSerSerSer 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13316 CCAAGTTAGAA.....GAAAGTACGACTTCTTAC 13344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1167 ProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGl 1183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13345 GAATCTACGATAAAGCGCTTA..... 13365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1183 uAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspL 1200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13366 AATGAAGAGATTACA..... 13380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1200 ysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGly 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13381 .....ACATTAAGAAGAAATAATGAACAAAGGCA 13413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1217 ...ArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAs 1232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13414 ATCCAGCAACAGATTACAGGTACATCTGCAATGAACAA.....AA 13454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1232 pAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgS 1249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13455 TGACTTATCAACATAGTTGAGTCTATGAAA..... 13485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 erArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpPro 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13486 .....AAGTCTTTTGAAGAAGATAAAATCAAAATTCATC 13518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1266 GluGluLys.....ArgLeuLysGluAspIleLeuGluAsnGluAspGl 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13519 AAAGAAAAACCCCAAGAGTTAATGAAAAATACTCGAGGCCCAAGAAAG 13568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1280 uGlnAsnSerProPro.....LysLysGlyLysA 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13569 GCTAAATCAACCTTCCCAATATCAATATGGAGGAGATTAAAAAAAATGGG 13618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1290 rgGlyArgProProLysProLeuGlyGlyThrProLysGluGluPro 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13619 AATCTGAGCAGCAACAGAGATATCTCAAAAGATTCGCGAAGCTGAGGAA 13668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1307 ThrMetLys.....ThrSerLysLysGlySerLys... 1316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13669 GCCCTCAAAAGCGCAATACAGATTACCCACTGAGGAGAAAAATAAAGAT 13718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1317 .....LysLysSerGlyProProAlaProGluGluGluGluG 1330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13719 AATCGAACGAAGAAGAGGGAATTGGAAAAAGAGTTTGAAGAAAGGTTG 13768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1330 luGluArg.....GlnSerGly..... 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```















[illegible][illegible]



30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetA 46  
||||| : : : : :  
1944 AGAATGTCCTTAAGCAACGTTTAGTGACCTTTTAGAATCACAAATCA 1993  
46 spGlnAspSerGluGluLysGluLeu...TyrLeuAsnLeuAlaLeu 61  
: : : : : : : : : : :  
1994 ACCGCTCTAAAGAGAATGAACAGATTAAGAGAGTTGAACACTGCAAG 2043  
62 HisLeuAlaSerAspPheLeuLysHisProGly..... 73  
: : : : : : : : : : :  
2044 GTAATAGCAGATGATTCAAAAAACAACCTCCGAAATGAAGACTTACT 2093  
74 LysAspValArgLeuLeu.....ValAlaCysCysLeuAlaAspI 87  
: : : : : : : : : : :  
2094 TAAGCAATTCAGTTAAGCAAGAAAAATAGCACAAATCCGAAAGAAT 2143  
87 lePheArgIle.....TyrAlaProGluAlaProTyrThr 98  
: : : : : : : : : : :  
2144 GTCTACGTTTATCCTCTATACTGACGAGCAGATGAAGATGAATGA 2193  
99 SerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 115  
||| : : : : : : : : : : :  
2194 TTATCTGCAAAATCTAGTCTGATTTATATCTTCTGAAGAAACAATTA 2243  
115 sgLysLeuGluAspThrLys.....SerProGlnPheAsnArgTyrP 129  
: : : : : : : : : : :  
2244 TAAAGAAAGCGCTACCAAGAACATCTTCAAAATCAAAATGAACATCA 2293  
129 heTyrLeuLeuGluAsn...IleAlaTrpValLysSerTyrAsnIleCys 144  
: : : : : : : : : : :  
2294 TCGTAGAGTTGGAACATAAGTGCCTCATATAAATCTTCAAA..... 2337  
145 PheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 161  
||| ||| ||| : : : : :  
2338 ...GAAGAACTGACATGTGGAAACGAATGAATACCGTGCATTTGTT 2384  
161 uPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetV 178  
: : : : : : : : : : :  
2385 ACTAGAGCATACATCGACGAGAGAATGCAAGGTAAAGGAATTAAGT 2434  
178 alaAspLeuMetSerSerIleIleCysGluGlyAsp.....ThrValSer 192  
: : : : : : : : : : :  
2435 CCAAAACCAAAAGCTAGTGGAAATGAAATGATCTTCAAACTTTAACT 2484  
193 GlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLys 209  
: : : : : : : : : : :  
2485 AAACACGCTCGAT.....CTATGCGCTCAAAATCAATACATCTTTAAT 2528  
209 nLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLysArgThrA 226  
||| : : : : : : : : : : :  
2529 TACCAATCTGTTCTTAATGACTCGAAGGACCTTACGTAAGGAAGAA 2578  
226 leGlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMet 242  
||| : : : : : : : : : : :  
2579 TTCAA.....TTTATTCAAAACATTATGCGAG 2604  
243 LeuGlyLysThrSerIle.....SerAspLeuSerGluHisValPheAs 257  
: : : : : : : : : : :  
2605 GAAGACGATAGTACTATCACAAATCTGACTTCAAAAAGTCGTAACTGA 2654  
257 pLeuIleLeuGluLeuTyrAsnIle.....AspSerHisL 269  
: : : : : : : : : : :  
2655 AAGACTAGTTGAATTCAAAACATATTATTCAATTAAGAAAAAATGCGAG 2704  
269 euLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAsp 285  
||| : : : : : : : : : : :  
2705 AACTTTTGAAGTAGTAAGAAATTCGCGGATAAGTTGGAATCGAAAGAA 2754  
286 AsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAl 302  
: : : : : : : : : : :  
2755 AGAATCTTAACAAGCTTCTCAGAAATC..... 2784

302 alYsAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrL 319  
: : : : : : : : : : :  
2785 ...GAAAGTGAACACAGTAAATGAGGCTAAA..... 2811  
319 euGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335  
: : : : : : : : : : :  
2812 .....GAGGCTATAATAACTTTAAGAGTGAAGAAATGGAT 2847  
336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuTh 352  
: : : : : : : : : : :  
2848 CTGAATCAAGAAAT.....GAGGAACACAGAAAGAGCTTGA 2885  
352 rGluTyrLeuLysValArgSerHisAspProGluAlaIleArgHisA 369  
||| : : : : : : : : : : :  
2886 AGAA...TTGAAACT...TCTGTTCCCAACGAAGATCGTTCATACAGCA 2929  
369 spValIleValSerIleValThrAlaAlaLysLysAspIle...LeuLeu 384  
: : : : : : : : : : :  
2930 ATGTAACATATAAACAGTAAACCGAAACTAAGAGAGACTTCAAACTCAA 2979  
385 ValAsnAsp.....HisLeuLeuAsnPheValArgGluArgThrLe 398  
||| : : : : : : : : : : :  
2980 GTACAGACTTCCAAACTCGTATCTCGCAAAATTAAGGAGTCTACT... 3027  
398 uAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnI 415  
: : : : : : : : : : :  
3028 GAAATATGTCACCTTTTAAACAAGGAG.....ATACAGGACC 3064  
415 leTyr...LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAla 430  
: : : : : : : : : : :  
3065 TGATGACAGCAAGCAGCATATCCATTAAGCTTGGAAAGGAAATCA 3114  
431 LysGlnIle...AlaTrpIleLysAspLysAlaHisIleTyrGln 446  
: : : : : : : : : : :  
3115 TCGCAATATTGGCAGAGGAGGATTTAACTACTT.....TC 3152  
446 nAsnSerIleAspAspArgLeuValGluArgIlePheAlaGlnTyrM 463  
: : : : : : : : : : :  
3153 GAATACGTTAGT..... 3165  
463 etValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 479  
: : : : : : : : : : :  
3166 .....CTAACTAAAGCTGAGACGACCACTCGCAAAAGGTTGATTAT 3210  
480 LeuTyrAlaThrLeu.....AspLeuAsnAlaValLysAlaLeuAs 493  
||| : : : : : : : : : : :  
3211 TTACAGAACTACTATTTTAAACAAGATTCCTCAAAACACAGACACTTAA 3260  
493 nGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 510  
||| : : : : : : : : : : :  
3261 TGAATACGTTTCTCTAAATCTAAGCTTAAGCATTTGTTGAAACAGAAATAT 3310  
510 euAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526  
||| : : : : : : : : : : :  
3311 TGAACCTG.....AAGAGAACACAGAAATTAAGAGTTTCAATTA 3348  
527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaG 543  
: : : : : : : : : : :  
3349 GAAAGAACTTCAAAACAGACTGAATAACTCTCCCTCGAAAG..... 3393  
543 nAspPheMetLysLysPheThrGlnValLeuGluAspGluLysIleA 560  
||| : : : : : : : : : : :  
3394 GACAGTTTACGATCATGGTAACCTCAATTAACAACTTTTACAAAAGGAGC 3442  
560 rGlyGlnLeuGluValLeuValSerProThrCysSerCysLys..... 574  
||| : : : : : : : : : : :  
3443 GTGAAGATCTATTGGAAGAGACTAGGAATCATGTCAAAAGAAATAGAT 3492  
575 GlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro... 590  
: : : : : : : : : : :  
3493 GAACITGAAGATGCTCTCAGCGAACTTAAAGGAACTTCTCAAAAGA 3542  
591 .....LysGlnProThrAsnProPheLeuGluMetI 601



4855 ATCCAGCAACAGTACACAGCTACATCTGCAAAATCAACAA.....AA 4895  
1232 pAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgS 1249  
4896 TGACTTATCTAACATAGTTGAGTCTATGAAA..... 4926  
1249 erArgLysArgGlyHisThrAlaSerGlySerAspGluGlnGlnTrpPro 1265  
4927 .....AAGTCITTTTGAAGAAGATAAAATCAATTCATC 4959  
1266 GluGluLys.....ArgLeuLysGluAspIleLeuGluAsnGluaspG1 1280  
4960 AAGAAAAACCCCAAGAGTAAATGAAAAAATCTCGAGGCCCAAGAAA 5009  
1280 uGlnAsnSerProPro.....LysLysGlyLysA 1290  
5010 GCTAATATCAACCTTCCATATCAATATGAGGAGAGATTAAAAAATGAG 5059  
1290 rGlyArgProProLysProLeuGlyGlyThrProLysGluGluPro 1306  
5060 AATCTGAGCAGCAACAGGAAGTATCTCAAAAGATTCGCGAAGCTGAGAA 5109  
1307 ThrMetLys.....ThrSerLysLysGlySerLys... 1316  
5110 GCCTCAAAAGCAATACAGATACCACTGAGGAGAAAATAAATAGAT 5159  
1317 .....LysLysSerGlyProProAlaProGluGluGluGluGluG 1330  
5160 AATCGAAGAAAGAGGAGGAATGGAAGAGTTGGAAGAAAGATTG 5209  
1330 luGluArg.....GlnSerGly..... 1335  
5210 AGGAGAGATAAATCAATGACCAATCTCGAGAAATAGACGTGGCTT 5259  
1336 .....AsnThrGluGlnLysSerLysGlnH1 1346  
5260 CGAAAACAGCTAGAACGTAAGTTCAGAGAAACAAAGGAATGGA 5309  
1346 sArgValSerArgArgAlaGlnGlnArgAlaGluSer....ProGluSerS 1362  
5310 CGAGTATACAAAAATTAATCAAGAAAGTCAAAAGATGTACCACATCAA 5359  
1362 erAlaIleGluSerThrGln 1368  
5360 GTCATATCTCAGATGAGAA 5379  
seq\_name: gb\_p14:SCYKR095W

seq\_documentation\_block:  
LOCUS SCYKR095W 8050 bp DNA PLN 18-MAR-1998  
DEFINITION S.cerevisiae chromosome XI reading frame ORF YKR095W.  
ACCESSION Z28320 Y13137  
VERSION Z28320.1 GI:486586  
KEYWORDS  
SOURCE Saccharomyces cerevisiae.  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
REFERENCE 1 (bases 1 to 8050)  
AUTHORS Baladron,V., Ballesta,J.P.G., Bou,G., del Rey,F., Esteban,P.F.,  
Garcia-Cantalejo,J.M., Garcia-Ramirez,J.J., Gonzalez,A.,  
Jimenez,A., Revuelta,J.L. and Santos,M.A.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 8050)  
AUTHORS MIPS.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-1994) Data collected by MIPS on behalf of the  
European yeast chromosome XI sequencing project. MIPS at the  
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152  
Martinsried, FRG; E-mail: Mewes@mpmic.mips.biochem.mpg.de  
FEATURES  
source  
1..8050  
/organism="Saccharomyces cerevisiae"

/db\_xref="taxon:4932"  
/chromosome="XI"  
complement(1..8)  
/gene="RPL40B"  
complement(1..8)  
/gene="RPL40B"  
/number=1  
complement(<1..8)  
/gene="RPL40B"  
/note="ORF YKR094C"  
/codon\_start=1  
/protein\_id="CAA82175.1"  
/db\_xref="GI:4388588"  
/translation="MQ"  
1063..6690  
/gene="MLP1"  
/note="MLP1"  
/note="ORF YKR095W"  
/codon\_start=1  
/protein\_id="CAA82174.1"  
/db\_xref="GI:486587"  
/db\_xref="MSDHDTPMESIQNGENSERLNAISFFGCSLEQVKSFGDGVVK  
HLNDKLOFNEKLSNKLVTYFDELKASSLKIDGLKTEMENVIRENDKIRERNDT  
FKPESVENEKMLKLSSELEFKVRLDLDTEKKETQSNOQRTKLILDERLKEIYRV  
ENNRSECKLRSTIMDETQOQYITNDLNSRTELEKTOELTLLOSNDWLEKEL  
RSKQYLSYRQTDKVJIDINELNRLNDFOMERTNDVLKQKNLSLSQLEKLL  
EIKGLSDLSNEKQFSAEMSLKQRLVDLSQNAVEELNSRELNTAKVIADDSK  
KQTPENEDLLKELQTLKELAQCEKCLSLSDIDEADENENLSAKSSDIFLKQ  
LKERRTEHLQNOJETFIVELEHKVPIINSFKERTDMLNENNAALLIERTSEKN  
AKVKELNAKNOLKEVENDLQTLTKQRDLQCRQIOYLLITNSVSDSKGPKRKEIQF  
IQNTMQEDDSITITESDSQKVTYRLEVEFKNIQLQEKNAELLKVYRNLDKLESEKK  
TQKQKIESETVNEAKGAILTLKSEKMDLESRIEELQELKELKTSVPNEASISNV  
TIKQITETKRDLESQVQDLQTRISQITRESTENMSLNKEIQDLYDSKSDISIKLGE  
KSSRLAEERFKLSNTLDTKAENDQRLKRFYDQNTILKQDSKTHETLNEYVSKS  
KLSIVTELLNLKEQKLVHLEKNLQELNKLSPKSDSLRMVTLQTLQKRVSDLL  
EETRKSCKDKLEDELSELKQKTHETQKDHHTKOLEEDNSNIENYQNKIEALKDYE  
SVITSVDSQTDIEKLOYKVKSLEKEIEEDKIRLHTYVMDETINDDSLRLEKSKI  
NLTDAYSQIKYKDIYETSSLOQTNKLDSEFKDFTNQIKNLDEKTSLEDKISLL  
KEQMFNLNDELQKQWKEKADFKRISILQNNKEVEAVKSEYKSLKIQNDLD  
QOTIYANTAQNNYEQELQHADVSKTISLREQHLYKQGVKTLNSRDLQENALKEN  
EKSWSQKESLEQDLNSRIEDLSQNKLLYDQIYITAAKQVNNSTNGPGLNI  
LITLRERDILDTKVTVARDAKMRQKISLMDVDELQDARTKLDNSRVKEKNSSTIQ  
QHDDIMEKLNOLLNRESNITLRNELENNNNKKKLOSELQDLKQNVAPIESELTKK  
YSMOEKEQLKAEVHRWKRSSQDILEKHEQLKSDYKLESEIENLKELENKER  
QGAEEERFNLRQQRKLTSLQSDSLTEQVNSLRDAKNVLENSSEANRIEEL  
QNAKVAQNNQLEAIRKQEDAEKASRELQAKLEESTTSYESTINGLNEEITTLKEERL  
EKQRIQOQLQATSANEQNDLSNIVESMKKSFEDKIKFKEKTOEVNEKILEAQERL  
NOPSINMEETKKWESEHQEVQKIREAEALAKRIRLPTEEKINKTIERKKEPELE  
KEFEKVEERIKSMEQSGEIDVLRKOLEAKVQKQKLENEYKQKLOPELQKVPHESS  
HISDDERKLAERIESLRBEFNELQAIKKKSFDEGQQAQMMKTLERKLAKESQ  
LSETKQSAESPSPKSVNNYONPLGLPRKIEENSFPNPLSGEKLKLLKNSKSSGGF  
NPFTSPSPKHLQNDKRESLANKTDPHTHLEPSFNIPASRGLTSSSTLSTDNDE  
ELTSNNPAQKDSNRNVQSEDETEKKKEGPKVKGAEIETEQTCKSNKRPTIDEVGLKND  
EDDTTENINESKKIKTEDEEKEKTDKVDNDSI"  
BASE COUNT 3194 a 1321 c 1540 g 1995 t  
ORIGIN  
alignment\_scores:  
Quality: 235.50 Length: 1540  
Ratio: 0.325 Gaps: 72  
Percent Similarity: 47.078 Percent Identity: 19.545  
alignment\_block:  
US-09-512-581-2 x SCYKR095W ..  
Align seg 1/1 to: SCYKR095W from: 1 to: 8050  
19 valysGluIleSerAspLysIleSerLysGlu..... 29  
:::||||| :::::|||||  
1894 ATCAAGGGTCTATCCGACTCCCTAAATTCGAAAAAGCAAGAAATTTCTGTC 1943







1763 TCGTAGAGTTGGAAACATAAAAGTCCCATTTATAAACTCTTTCAAA..... 1806  
145 PheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLe 161  
1807 ...GAAAGAACTGCATGTTGGAAACAGAAATGAATAACGCTCATGTT 1853  
161 uPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetV 178  
1854 ACTAGACATACATCGAACGAGAGAAATGCAAGGTTAAGGAATTAATG 1903  
178 alAspLeuMetSerSerIleCysGluGlyAsp.....ThrValSer 192  
1904 CCAAAACCAAAAGCTAGTGGAAATGTGAAATGATCTTCAAACCTTAACT 1953  
193 GlnGluLeuLeuAspThrValLeuValAsnLeuValProLalaHisLys 209  
1954 AACAAGCTCTCGAT.....CTATGCCGTCAAAATACAATACCTTTTAAT 1997  
209 nLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrA 226  
1998 TACCAATTCGTGTTCTTAATGACTCGAAGGACCCCTACGTAAGAGAAA 2047  
226 laGlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMet 242  
2048 TTCAA.....TTTATTCAAAACATTATGCAG 2073  
243 LeuGlyLysThrSerIle.....SerAspLeuSerGluHisValPheAs 257  
2074 GAAGACGATAGTACTATACAGAACTGACTCTCAAAAAGTCGTAACGTA 2123  
257 pLeuIleLeuGluLeuTyrAsnIle.....AspSerHisL 269  
2124 AAGACTAGTTGAATTCAAAACATATTCAATTACAGAAAAAATGCAG 2173  
269 euLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAsp 285  
2174 AACTTTTGAAGTAGTAAGAACTTACGCCGATAAGTTGGAATCGAAAGAA 2223  
286 AsnGluCluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAl 302  
2224 AAGAATCTAAACAAAGCTCTTCAGAAATC..... 2253  
302 alYsAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrL 319  
2254 ...GAAAGTGAACAGCTAAATGAGGCTAAA..... 2280  
319 euGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335  
2281 .....GAGGCTATAATAACTTTAAAGAGTGAAGAAATGGAT 2316  
336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuTh 352  
2317 CTAGAATCAAGAAT.....GAGGAACCTACAGAAAGAGCTTGA 2354  
352 rGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369  
2355 AGAA...TTGAAAACT...TCTGTTCCCAACGAGATGCGCTACACGCA 2398  
369 spValIleValSerIleValThrAlaAlaLysLysAspIle...LeuLeu 384  
2399 ATGTAACATATAAACAGTTAACCGAAACTTAAGAGACCTCGAATCTCAA 2448  
385 ValAsnAsp.....HisLeuLeuAsnPheValArgGluArgThrLe 398  
2449 GTACAAGACTTGCAAACTCGTATCTCGCAAAATTACTAGGGAGCTACT.. 2496  
398 uAspLysArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnI 415  
2497 .GAAATATGCTACTTTTAAACAGGAG.....ATACAGGACC 2533  
415 leTyr...LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAla 430  
2534 TGTATGACGAAGAGGACATATCCATTAAAGCTTGGAAAGGAAAAATCA 2583

431 LysGlnIle...AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGl 446  
2584 TCAGAAATATTGCAGAGGAACGATTTAAACTACTT.....TC 2621  
446 nasnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrM 463  
2622 GAATACGTTAGAT..... 2634  
463 etValProHisAsnLeuGluThrThrGluArgMetLysCysLysLeuTyrTyr 479  
2635 .....CTAATAAGCTGAGACGACCAACTGCGCAAAAGGTTTGATTAT 2679  
480 LeuTyrAlaThrLeu.....AspLeuAsnAlaValLysAlaLeuAs 493  
2680 TTACAGAACTACTATTTAAACAAGATTCCAAAACACACAGACACTTAA 2729  
493 nGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuL 510  
2730 TGAATACGTTTCTCTGTAATCTAAGTTAAGCAITTTGTTGAAACAGAAATAT 2779  
510 euAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526  
2780 TGAACCTG.....AAAGAGACAGAGAAATTAAGAGTTTCATTTA 2817  
527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGl 543  
2818 GAAAGAACTTGAACAAGAACTGAATAAACTCTCCCTCGAAAG..... 2862  
543 nAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleA 560  
2863 .GACAGTTTACGCATCATGCTAACTCAATTACAACTTTACAAAAGCAGC 2911  
560 rgLysGlnLeuGluValLeuValSerProThrCysSerCysLys..... 574  
2912 GTGAAGATCTATTGGAAGAGACTAGGAAATCATGTCAAAAGAAATAGAT 2961  
575 GlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro... 590  
2962 GAATCTGAAGATGCTCTCAGCACTTTAAAGAGAACTTCTCAAAAGA 3011  
591 .....LysGlnProThrAsnProPheLeuGluMetI 601  
3012 CCATCATATCAACAGCTGGAAGAGACAACTTCAAAATATAGAAATGGT 3061  
601 leLysPheLeuLeuGluArgIleAlaProValHis..... 612  
3062 ACCAAATAAAATCGAAGCTTTGAAGAAAGATTATGAATCAGTAATAACT 3111  
613 .....IleAspThrGluSerIleSerAlaLeuIleLy 623  
3112 TCTGTAGATAGTAAGCAACTGACATTGAGAAATTTACAATATAAGTCAA 3161  
623 sGlnValAsnLysSerIleAsp.....GlyThrAlaA 634  
3162 ATCACTAGAAAAGAAATCGAGGAGGACAACTTCGTTTACATACTTATA 3211  
634 spAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650  
3212 ATGTTATGATGAAACAAT...AACGATGATTCCCTACGCAAGGAGTTG 3258  
651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667  
3259 GAAAAATCCAAAGATT...AACTTAACATGAT.....GCTTATTCACAAAT 3299  
667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684  
3300 CAAAGATACAGGATCTCTAC..... 3321  
684 alaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700  
3322 ....GAGACTACTCTCAGTCTTTGCGCAACAGAAATCTTAATTTGGAT 3366







[illegible]

6147	TTcATATGAAGCAGTGCCTGGTATGAAGATG.....	.....TCT	6181
990	TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa	1006	
1111	:   :		
6182	TACCAGGTACAACAGGCAATCAACACATGCCTAAAGATCCT.....	.....	6223
1006	LlysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp	1023	
1111	:  :   :  :   :  :   :  :   :	:  :   :  :   :	
6224	...GTAGGGGTTCAGACAA...GAGCAGTCCTCTAGCGCTTGTGT	6266	
1023	heValLeuGluIleuMetAlaLysAsnGluuAsnSerHisAlaPhe	1039	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6267	CACACCTTTACTCCATGATCGTGGAAACCGCAACACAGACGCTTT	6316	
1040	IleArgLysMetValGluuAsnIleLysGlnThrLysAspAlaGlnGlyPr	1056	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6317	CTTATTCTTTACTCAACCTCTTT.....	.....GA	6342
1056	oAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaM	1073	
1111	:   :  :   :  :   :  :   :	:   :  :   :	
6343	TGACACAGCAAAA.....	.....ACAGACGTGACTA	6368
1073	eTAsnIle...IleMetSerLysSerThrThrTyrSerLeuGluSerPro	1088	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6369	TGCTCTTTGATATACAGACAATCTAGCCTGTTTCCATACCACACACAG	6418	
1089	LysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPh	1105	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6419	GAAGAGCGCTGTTT.....	.....	6433
1105	eSerAsnThrLysAsnTyrLeuProGluMetLysSerPheThrPr	1122	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6433	.....	.....	6433
1122	roGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu	1138	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6434	.....ATAATGCATCATATACACATTCACATC	6460	
1139	SerSerAlaGly.....LysGlnSerGlnThrLys	1148	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6461	TCAGTTCTGGTAGTAACCTACTGCAGTCATTCAAGGAGTCTATGGTAA	6510	
1148	sSerArgArgMetGluThrValSerAsnAlaSer.....	1159	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6511	GGACAAAGGAAAGACAGACAAATCATCACCTAGTAAGGAAATAGACTCA	6560	
1160	.....SerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeu	1174	
1111	:  :   :  :   :  :   :  :   :	:  :   :  :   :	
6561	CGCAGACTGAGAAAGATTTCCAGCCTCGGAAGTCACGGAACAGCTGTA	6610	
1175	AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSe	1191	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6611	GATTTCAGATTCAGATTTCAGATTCAGAACGCGATATAATTCAGTATGAA	6660	
1191	rProLeuProGlyLysSerAspLysArgAspSerAspLeuValA	1208	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6661	ATGTTTCCGCAAAATTCAGCTCCTTTAATCGAATTCGAATGTC.....	6706	
1208	rgSerGluLeuGluLysProArgGlyArgLysLysThrProValThrGlu	1224	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6707	.....TCCAGGGTATTTTATTACTTCTCATGTTAAAA	6739	
1225	GlnGluGluLys.....LeuGlyMetAspAspLeuThrLysLeuValG	1239	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6740	CAACATTGGAAGATCTTTGTGGATTTTCTGAT.....AGTAAATTC	6783	
1239	nGluGlnLysProLysGlySerGlnArg.....SerA	1250	
1111	:   :  :   :  :   :  :   :  :   :	:  :   :  :   :	
6784	GAAGTACTCTCCATCTGATCTGCAGAAAGTATATGATAAGCGATAAAC	6833	
1250	rgLysArgGly.....	1253	
1111	:   :  :   :  :   :  :   :	:  :   :  :   :	
6834	GAATAACAGGAGTTCATTTCATCCAAACAAACACTGGACATTCCTGCG	6883	



```

291 lnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu... 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3708 ACTTCTCTCCATCTCTTAAATAATGTGGTAGTAAGCAAGGTGAAGAA 3757
307 .....LeuAlaSerGI 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3758 GATTACAGACCAGCTGTTTGAATAATTTTGTTCAGACCTTCTTTCAACAGT 3807
310 nAsnLysProLeuTyrGlnCys.....TyrLeuGlyA 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3808 CATAAGCCGTGAATGCCAGCTCTCAACTACTCTTACTCTTGTAGTTAGGGA 3857
321 rgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAla 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3858 GA..... 3859
338 SerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTy 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3860 .....CTGTGTGTTTCATCATGTTCAAGTAAACAAGTCAACAGAGATGGC 3900
354 rLeuLysValArgSerHisAsp.....ProGluGluAlaIleA 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3901 TTTAAGAGTGGCATCTCTTCACTTACCTTGAACTGTTGCTGCACGCGTAA 3950
367 rgHisAspValIleValSer..... 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3951 GAAAGATGCTGTTACAAGCAAAATGGATCAAGGATCTATAGAAGCAATT 4000
374 .....IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHI 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4001 TTAAACAGGTTTCAGGAGGGGAAGATGAATCCACAACATTTACAAAAGC 4050
388 sLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTyrArgVal. 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4051 ATTGCTTTGATTACTTGGATGAACAACACTGAGACTGATCTTCACTAGTGT 4100
405 .....ArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4101 TTTCCTCGTAAA.....TTCTATATAGCCCAAGTGGTTTCGAGACACA 4141
420 AlaLeu.....GlnSerAl 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4142 ACTCTGGAACACAGAAAGCAATGAATCACAATAAGATGAAGAATCNYC 4191
424 aAlaGlyLysAspAlaAlaLysGlnIle..... 433
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4192 TGAAGGAACACATCATGCAAGGAAATTTGAGACAACCTGGCCAAATTATGC 4241
434 .....AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSer 448
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4242 ATCGAGCTGAAACACCGAAAAAGTTTCTTAGAAGCATTTATCAAAACCACA 4291
449 .....IleAs 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4292 CCTTCTCAGTTTAGCACATTAAGATGAACCTGATCATGTGGACTATGA 4341
450 pAspArgLeuLeuValGluArgIle.....PheAlaG 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4342 TGATGCTGTGCTTATGTTTCGATACTTGGCTCCATGAGGCCGTTTGCCC 4391
461 lnTyrMetValProHisAsnLeuGlu..... 469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4392 AGAGCTTTGATATTTATTTTGACACAGATCCTACGAGTTCTTGTGTGAANAAT 4441
470 .....ThrThrGluArgMetLysCysLeuTyr..... 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4442 GCAATTGCTGTTTCGACAAAAAGCCATGAAGTGTGTTGCTGAGGTGTTGCC 4491
479 .....TyrLeuTyrAlaThrLeuAspLeuAsn..... 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4492 TGTAGACCCAGTATTCTAGCAAGGCTTGATATGCAACGAGGTGTTTCATG 4541

```

```

488 .....AlaValLysAla 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4542 GAGCATTTGATGGATAATTCGACTAGTGTCCGAGAAGCAGCAGTAGAATTA 4591
492 LeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAs 508
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4592 CTAGGTCGATTTTGTCTTGTGCACTCAGCTTGTGTAACAGATATTATGA 4641
508 pLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaI 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4642 TATGCTG...ATTGAAAGAATATTGGATACTGTGTATCATGTCAGG.... 4684
525 lePheSerLysValMetValIleThrArgAsnLeu.....Pro 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4685 .....AAAAGATTAATAAGATTTCTCAGAGACATTTGTATTGAACAACCA 4729
538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGI 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4730 ACATTTCCAAAAATCACAAAATGTGTGTAATAATGATTCGCAGAGTCAA 4779
554 uAspAspGluLysIleArgLysGlnLeuGluValArgGluIleThrLysLys 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4780 TGATGAAGAGGCGCATTAAGAAA..... 4801
571 ysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4802 .....TTAGTAATGAACATTTCCAGAAACTC 4828
588 GlyAsnProLysGlnProThrAsnPro..... 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4829 TGGTTTACTCCAACTCCACAATGACAAAGAAGCAATGACAAGGAAAT 4878
596 ..... 596
4879 TTTAAACATTACCGATGTGGTTCGACATGCAGAGATACTGGATATGACT 4928
597 ...PheLeuGluMetIleLysPheLeuLeuGlu.....Arg 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4929 GGTTCGAGCAACTGCTTCAAACTTCTTGAAGTCTGAAGAGAGATTCTCTCA 4978
608 IleAlaProVal...HisIleAspThrGluSerIleSerAlaLeuIleLy 623
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4979 TATAAACCTGTGAAGAAAGCTTGTACTCAACTTGTGATAACCTAGTTGA 5028
623 sGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyV 640
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5029 GCACATTTCTTAATATGAGGAATCTCTAGCTGACTCTGACAATAAAGGTG 5078
640 aProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeu 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5079 TGAATTTCTGGAAGATTGGTA...GTTGCATAACCACTTTGTTCTTATTC 5125
657 SerPheThrHisPro...IleSerPheHisSerAlaGluThrPheGluSe 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5126 AGCAAAATAGACCCCGCTCATGGTTAAACATGCAATGACTATGCAACC 5175
672 rLeuLeuAla.....CysLeuLysMetAsp..... 681
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5176 ATACCTTACCACTAAATGTAGTAGCAAAATGATTTTCATGGTTATCTGCA 5225
682 .....GluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5226 ATGTTCAAAAAATCCTAGAGCTAGTTGTACCACCTGATGGAGCATCCAAGT 5275
697 .....SerLysIleGluGluAspPheProHis.....IleAr 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5276 GAAACTTTTCTGCCACTATTGAGGAAGATCTAATGAAGCTCATCATCAA 5325
707 gSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProA 724
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5326 ATATGGCATGACTGTAGTGAACAT.....5350
724 rgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740

```











450 pAspArgLeuLeuValGluArgIle.....PheAlaG 461  
 432 TGATGCTTGGCTGTGATATGTTTCGATACTTGGCTCCATGAGCCGTTGGCC 4391  
 461 InTyMetValProHisAsnLeuGlu..... 469  
 4392 AGAGCTTGTATTTATTTATGTGACACAGATCCTACAGTCTTGTGTGAAAT 4441  
 470 .....ThrThrGluArgMetLysCysLeuTyr..... 478  
 4442 GCAATTGCTGTTCGAACAAGCCATCAAGTGTGTGCTGAGCTTGTTCG 4491  
 479 .....TyrLeuTyrAlaThrLeuAspLeuAsn..... 487  
 4492 TGTAGACCCAGTATTTCTAGCAAGGCTTGATATGCAACAGAGGTGTTCATG 4541  
 488 .....AlaValLysAla 491  
 4542 GACGATTTGATGGATAATTCGACTAGTGTCCGAGAGCAGCAGTAGAATTA 4591  
 492 LeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAs 508  
 4592 CTAGTGCATTTGCTTGTTCGACCTCAGCTTGTCTGAACAGTATATGA 4641  
 508 pLeuLeuAspLeuLeuLysGlnProLysThrAspAlaSerValLysAlaI 525  
 4642 TATGCTG...ATTGAAGAAATATTGGATCTGTTATCAGTGTCTCAGG.... 4684  
 525 IePheSerLysValMetValIleThrArgAsnLeu.....Pro 537  
 4685 .....AAAAGAGTATAAAGATTCTCAGACACATTTGTTATGAACAACCA 4729  
 538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuG 554  
 4730 ACATTTCCAAANAATCAGAAATGTGTGTAANAATGATTCGACAGATCAA 4779  
 554 uAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrC 571  
 4780 TCATGAAGAGGCATTAAAGAA..... 4801  
 571 ysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu 587  
 4802 .....TTAGTAAATGAACATTCCAGAACTC 4828  
 588 GlyAsnProLysGlnProThrAsnPro..... 596  
 4829 TGGTTTACTCCAACTCCACACATGCAAGAAGCAATGACAGGAAAT 4878  
 596 ..... 596  
 4879 TTTAAACATTACCGATGTGGTTGCAGCATGCAGAGATCTGGATATGACT 4928  
 597 ..PheLeuGluMetIleLysPheLeuLeuGlu.....Arg 607  
 4929 GGTGTTCAGCACTGCTTCCAAACACTGTGTGAAGTCTGAAGAGGATTCCTCA 4978  
 608 IleAlaProVal...HisIleAspThrGluSerIleSerAlaLeuIleL 623  
 4979 TATAAACCTGTGAGAAGAGCTGTGTACTCACTTGTGTGATAACTAGTATGA 5028  
 623 sGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGly 640  
 5029 GCACATCTTAAATATGAGGAATCTCTAGCTGACTCTGACAATAAAGGTG 5078  
 640 alProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeu 656  
 5079 TGAATTCTCGAAGATTGGTA..GCTTGCATAAACCACTTTGTTCTTATTC 5125  
 657 SerPheThrHisPro...IleSerPheHisSerAlaGluThrPheGluSe 672  
 5126 AGCAAAAATAGACCCCACTCATGGTTAAACATGCAATGACTATGCAACC 5175  
 672 rIeuLeuAla.....CysLeuLysMetAspAsp..... 681

5176 ATACCTTTACCACTAAATGTTAGTACGCAAAATGATTTTCATGGTTATCTGCA 5225  
682 ....GluLysValAlaGluAlaLeuGlnInIlePheLysAsnThrGly 696  
5226 ATGTTGC AAAATCCTAGAGCTAGTTGTACCACCTGATGGAGCATCCAAGT 5275  
697 .....SerLysIleGluAspPheProHis.....IleArg 707  
5276 GAAACTTTTCTTGCACATATTGAGGAAGATCTAATGAAGCTCATCATCAA 5325  
707 gSerAlaIleLeuProValLeuHisHisLysSerLysLysGlyProPro 724  
5326 ATATGGCATGACTAGTGCACAT..... 5350  
724 rgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740  
5351 .....TGTTGAGCTGCTGTGGAGCTGTTCTTAATGATACATATGG 5433  
741 GluThrGln.....PheAlaGlnIlePheGly 749  
5384 GTGACACAAAATTTAAATTTGTGTGGGCTTGTTCATGATACATATGG 5433  
749 uProLeuHisLys.....SerLeuAspProSerAsnLeuGly 761  
5434 TGCCATTTCAAAATTA AAAAGTCAACACGACAGAGCCCAATAATACACTT 5483  
761 lu.....HisLeuIleThrProLeuValThrIleGly 771  
5484 CACTTCAACAAACAAACGACACTTCTTAGATCCCTTTTCCACCGTTGGA 5533  
772 HisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTr 788  
5534 GCATATGT.....CGCATTTT..... 5551  
788 pValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProG 805  
5552 .....GATTTTCATCGAAGATTTTAAAGGCCAACAGCAAGGTTAAACA 5594  
805 lLys.....LysThrThrLysLeuTrpVal.....ProAsp 815  
5595 TAAAGATTAAGTACTTGAACATTTGATGTATTTCACAAACACTCAGAT 5644  
816 GluGluValSerProGluThrMetValLysIle..... 826  
5645 GAAGAAGTACAAACAAAGCTATCATTTGGCTAGGATTTGCCTTTTATTC 5694  
827 .....GlnAlaIleLysMetMetValArgTrpL 836  
5695 GCATCCAAAGTCTAATGTTTCGAGCAAGAAGTGAAGAATCTATATAATA 5744  
836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852  
5745 TTTTCTGTGATGAAGCACTCCCTCAGTCAATTTAAAAATACAGGTGTTAAA 5794  
853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGly.. 868  
5795 AACCTCCAGACCTACCTACAGAAGAAGATACAGTATCCAGCAGCAGCA 5844  
869 .....LysIleSerLys.....ProAspMetSer 877  
5845 TAGAGACTGGAAGAAAGTTGCAAAACAGGACACTTAAAAAGAAATGGGTG 5894  
877 rgLeuArgLeuAlaAlaGlySerAlaIleValLysLeu..... 889  
5895 ATGTTTCCCTCAGGATGATGATGATTCATCATCATGACGCTTTATCTCAAACAG 5944  
890 AlaGlnGluProCysTyrHisGluIleIleThrLeuGluGluThrGlnLe 906  
5945 GTGCTTGAGGCATTTTTTCACACCCAGTCAGTACGCGCACCTTTGCCCT 5994  
906 ucYsAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPhe 923











```
260 CAATCTCCTTATTATCCAGATGAAATTTTGACCTGAGAAATTTATT 309
923 AlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTy 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 TTCACCAAGCTTCAGAGCAGCTGCAACTTAAAGAGTTGCCAATTAGTTA 359
939 rMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlyArgArgA 956
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 TTATCCTTTATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 409
956 laHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTy 972
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 CTAAGCG.....TCTATATGG 426
973 LeuLysGlnHisAlaValSerGluLysLeuLeuSerLeuLeuProGln 989
' : : : : : : : : : : : : : : : : : : : : : : : : :
427 ATACGTTCTCAGGTTGCTTTTCCAGAAACCCATGATTTTACCATTGGA 476
989 uTyValValProTyThrIleHisLeuLeuAlaHisAspProAspTyrv 1006
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 ATATGTAGCAACATATCTTATCCATCTCTCTCTCACCATCCAGACATAT 526
1006 allYsValGlnAspIleGluInLeuLysAspValLysGluCysLeuTyr 1022
: : : : : : : : : : : : : : : : : : : : : : : : :
527 CTTCGATAGATCAGAGAACTCGCTGATTTTATTGCATACATTCGCTTT 576
1023 PheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPh 1039
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 TAGGTTGACACTGTTCTTAATAGTAGGAAAGCTC.....CCAAT 614
1039 eLeArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyP 1056
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
615 CGTTTTCATCTAATCAACGCAATAAACAAGCTATGATGTT.....A 658
1056 roAspAspAlaLysMetAsnGluLysLeuTyThrValCysAspValAla 1072
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 TTGAAGATGGA.....ACAATTACATTTATGCTTAAGCGCATGGCT 702
1073 MetAsnIleLeuMetSerLysSer.....Thrh 1082
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 CAGAAATATTCAGCTGAAATCCAGAACTTTGGATGGATTTTGACTAC 752
1082 rTy.....SerLeuGluSerProLysAspProValLeuProAlaArgp 1097
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
753 GTATCCAAAGCAAAATTAATATCTTACGAAATTTTAAGACCAATTCCT 802
1097 hepThrGlnProAspLysAsnPheSerAsnThrLysAsnTyLeuPro 1113
: : : : : : : : : : : : : : : : : : : : : : : : :
803 CTATAGATGAGAAAAGAGAAATTTTCAAT.....AAATCTTCATAACG 846
1114 ProGluMetLysSerPheThrProGlyLysProLysThrThrAsnVa 1130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
847 CCTAAATGGAAGT.....CAGAT 866
1130 lLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnT 1147
: : : : : : : : : : : : : : : : : : : : : : : : :
867 TGAACATGCTATAAGACACACCTCTCATCTTTTGGCAACACAGCTACCA 916
1147 hTySerSerArgMetGluThrValSerAsnAlaSerSerSerAsn 1163
: : : : : : : : : : : : : : : : : : : : : : : : :
917 ATAAAGCAGCAATTTGAAGCAAAAGAGACATCATAGCTCAAGAGCGAT 966
1164 ProSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAs 1180
: : : : : : : : : : : : : : : : : : : : : : : : :
967 AAAAATCGCTAGAGAGGAAA.....AAGCAAAAAG 1001
1180 pHisSerGluAsnLysAspTyThrMetSerProLeuProGlyLysL 1197
: : : : : : : : : : : : : : : : : : : : : : : : :
1002 GAGAAAGCTAAATGACAGAAATCCTAATATTAGGACGTCGCCAGAACGAA 1051
1197 ySerAspLysArgAspAspSerAspLeuValArgSerGluLeuLys 1213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1052 GTAGCTCAAGATTTCAAGGATTCGTATAAATTTATTCGGAAGCACCGCTCT 1101
```

```
1214 ProArgGlyArgLysLysThrProValThrGluGlnGluLysLeuG1 1230
: : : : : : : : : : : : : : : : : : : : : : : : :
1102 .....TCAAGTGAGGAGATAAGCGAAGAGAGAGGAGGAATTAG 1139
```

```
1230 yMetAspAspLeuThrLysLeu 1237
: : : : : : : : : : : : : : : : : : : : : : : : :
1140 CGAAGAAGATTTTGTGATGAATA 1161
```

seq\_name: gb\_hgt20.AL353724

seq\_documentation\_block:

LOCUS AL353724 163642 bp DNA HTG 23-JAN-2001  
DEFINITION Homo sapiens chromosome 13 clone RP11-448113, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 9 unordered pieces.

ACCESSION AL353724

VERSION AL353724.3 GI:9930876

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 163642)

Burton, J.

Direct Submission

Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 27, 2000 this sequence version replaced gi:99265534.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA448113

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: pGAP4; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 159306 bases at least Q40

Consensus quality: 161143 bases at least Q30

Consensus quality: 162084 bases at least Q20

Insert size: 162842; sum-of-contigs

Insert size: 163676; 2.5% error; agarose-fp

Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality

coverage: 4.42x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 9 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 20261: contig of 20261 bp in length

\* 20262 20361: gap of 100 bp

\* 20362 38140: contig of 17779 bp in length

\* 38141 38240: gap of 100 bp

\* 38241 85018: contig of 46778 bp in length

\* 85019 85118: gap of 100 bp

\* 85119 93385: contig of 8267 bp in length

\* 93386 93485: gap of 100 bp

\* 93486 96470: contig of 2985 bp in length

\* 96471 96570: gap of 100 bp

\* 96571 137301: contig of 40731 bp in length

\* 137302 137401: gap of 100 bp

\* 137402 147035: contig of 9634 bp in length

\* 147036 147135: gap of 100 bp

\* 147136 149710: contig of 2575 bp in length

\* 149711 149810: gap of 100 bp

\* 149811 163642: contig of 13832 bp in length.







Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,  
 Ferrelira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,  
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K.,  
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
 Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 16, 2000 this sequence version replaced gi:6978239.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L6193  
 Center clone name: 343\_C\_9

----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 145534 bases at least Q40  
 Consensus quality: 151792 bases at least Q30  
 Consensus quality: 154567 bases at least Q20  
 Insert size: 194000; agarose-fp  
 Quality coverage: 157263; sum-of-contigs  
 Quality coverage: 3.2 in Q20 bases; agarose-fp  
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 38 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1187: contig of 1187 bp in length  
 1188 1287: gap of 100 bp  
 1288 2962: contig of 1675 bp in length  
 2963 3062: gap of 100 bp  
 3063 4458: contig of 1396 bp in length  
 4459 4558: gap of 100 bp  
 4559 5583: contig of 1025 bp in length  
 5584 5683: gap of 100 bp  
 5684 7420: contig of 1737 bp in length  
 7421 7520: gap of 100 bp  
 7521 9277: contig of 1757 bp in length  
 9278 9377: gap of 100 bp  
 9378 11138: contig of 1761 bp in length  
 11139 11238: gap of 100 bp  
 11239 13265: contig of 2027 bp in length  
 13266 13365: gap of 100 bp  
 13366 15440: contig of 2075 bp in length  
 15441 15540: gap of 100 bp  
 15541 17308: contig of 1768 bp in length  
 17309 17408: gap of 100 bp  
 17409 19894: contig of 2486 bp in length  
 19895 19994: gap of 100 bp

19995 21852: contig of 1858 bp in length  
 21853 21952: gap of 100 bp  
 21953 24275: contig of 2323 bp in length  
 24276 24375: gap of 100 bp  
 24376 27231: contig of 2856 bp in length  
 27232 27331: gap of 100 bp  
 27332 31148: contig of 3817 bp in length  
 31149 31248: gap of 100 bp  
 31249 33450: contig of 2202 bp in length  
 33451 33550: gap of 100 bp  
 33551 36754: contig of 3204 bp in length  
 36755 36854: gap of 100 bp  
 36855 39988: contig of 3134 bp in length  
 39989 40088: gap of 100 bp  
 40089 42367: contig of 2279 bp in length  
 42368 42467: gap of 100 bp  
 42468 45775: contig of 3308 bp in length  
 45776 45875: gap of 100 bp  
 45876 49987: contig of 4112 bp in length  
 49988 50087: gap of 100 bp  
 50088 53225: contig of 3138 bp in length  
 53226 53325: gap of 100 bp  
 53326 56622: contig of 3297 bp in length  
 56623 56722: gap of 100 bp  
 56723 59393: contig of 2671 bp in length  
 59394 59493: gap of 100 bp  
 59494 63369: contig of 3876 bp in length  
 63370 63469: gap of 100 bp  
 63470 68148: contig of 4679 bp in length  
 68149 68248: gap of 100 bp  
 68249 72051: contig of 3803 bp in length  
 72052 72151: gap of 100 bp  
 72152 78079: contig of 5928 bp in length  
 78080 78179: gap of 100 bp  
 78180 82389: contig of 4210 bp in length  
 82390 82489: gap of 100 bp  
 82490 87187: contig of 4698 bp in length  
 87188 87287: gap of 100 bp  
 87288 93791: contig of 6504 bp in length  
 93792 93891: gap of 100 bp  
 93892 99177: contig of 5286 bp in length  
 99178 99277: gap of 100 bp  
 99278 104756: contig of 5479 bp in length  
 104757 104856: gap of 100 bp  
 104857 112644: contig of 7788 bp in length  
 112645 112744: gap of 100 bp  
 112745 119589: contig of 6845 bp in length  
 119590 119689: gap of 100 bp  
 119690 130279: contig of 10590 bp in length  
 130280 130379: gap of 100 bp  
 130380 144451: contig of 14072 bp in length  
 144452 144551: gap of 100 bp  
 144552 160962: contig of 16411 bp in length.

FEATURES  
 Source

Location/Qualifiers  
 1..160962  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-343C9"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1..1187  
 /note="assembly\_fragment"  
 1288..2962  
 /note="assembly\_fragment"  
 3063..4458  
 /note="assembly\_fragment"  
 4559..5583  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right  
 5684..7420  
 /note="assembly\_fragment"

misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature



```
* 59861 61557: contig of 1697 bp in length
* 61558 61657: gap of 100 bp
* 61658 64020: contig of 2363 bp in length
* 64021 64120: gap of 100 bp
* 64121 65426: contig of 1306 bp in length
* 65427 65526: gap of 100 bp
* 65527 68141: contig of 2615 bp in length
* 68142 68241: gap of 100 bp
* 68242 70236: contig of 1995 bp in length
* 70237 70336: gap of 100 bp
* 70337 72882: contig of 2546 bp in length
* 72883 72982: gap of 100 bp
* 72983 75761: contig of 2779 bp in length
* 75762 75861: gap of 100 bp
* 75862 78586: contig of 2725 bp in length
* 78587 78686: gap of 100 bp
* 78687 81713: contig of 3027 bp in length
* 81714 81813: gap of 100 bp
* 81814 83813: contig of 2000 bp in length
* 83814 83913: gap of 100 bp
* 83914 87237: contig of 3324 bp in length
* 87238 87337: gap of 100 bp
* 87338 89353: contig of 2016 bp in length
* 89354 89453: gap of 100 bp
* 89454 91706: contig of 2253 bp in length
* 91707 91806: gap of 100 bp
* 91807 94574: contig of 2768 bp in length
* 94575 94674: gap of 100 bp
```

```
alignment_scores:
  Quality: 326.00      Length: 351
  Ratio: 1.743        Gaps: 15
  Percent Similarity: 53.276  Percent Identity: 33.048

alignment_block:
US-09-512-581-2 x AC090081 ..
Align seg 1/1 to: AC090081 from: 1 to: 167671
```

```
35 LeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluG1 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77205 TTACAGATGGTAGTGAACCTTTATGGATATGGACTCAGGACTCAAGA 77254

51 uGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhe 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77255 TGAANAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAAATCT 77304

68 heLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeu 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77305 TCCTCAGGACCCCAATAAAGATGTGGCTCTCCCTGTAGCATGTGTTG 77354

85 AlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAs 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77355 GCTGATATCTTCGTATCTATGCCCAAGAGCTCCATATACATCTCCCATGA 77404

101 pLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuG 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77405 TAAACTTAAGTAATAACAGCCTTTTACCTTTATTTATTTTTCATTTCA 77454

118 luAspThrLysSerProGlnPheAsnArgTyrPheTyrLeu..... 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77455 CCACGTGTCAGAGGTATACCTT.....TATTTTATTTTACCTTTTA 77498

132 .....LeuGluAsnIleAlaTrpVa 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77499 TGAACATAGGTTTCATATTAGTCTCTTCATCTTAAGAATATATTGTAAGT 77548

138 1..... 138
77549 TGCAAAATATTCAATTCAGTAGTATTTCACAAATCAGCTGGGTGCTTG 77598

139 .....LysSerTyrAsnIle 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
77599 GTTGTGTTGATTGAAGCATAGAAAGGTATAGTTCAAACACTAGTCCCTGTA 77648

144 CysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArg... 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77649 TGC.....TCAATCCCAATTTTCTGGCTAATTCATCACAA 77683

160 .....ThrLeuP 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77684 AAGCCAGATAGCTATGTTCCAGGGATAAATAGCAGATTAAATGATGACTTT 77733

162 heSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetVal 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77734 TTAATAATA.....GCATACCAGAATTTTACATATTTCATGACT 77774

179 .....AspLeuMetSerSerIleI1 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77775 ATTTTCCCTATAGTTATTATCAGCATGCAAAACATATATGCTCATCTTGGT 77824

185 eCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal..LeuVal 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77825 TGTCAA.....GTGTTTGTGGAACTCTTAACATTTGATCTCTTGGG 77865

202 AsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAl 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77866 GAGATTTTGTATTTTGAATAATTTTCAATCTTGGCAATTTTAAATCCTGTA 77915

218 aLysAlaLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrT 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77916 GACTGCTTTCTTTAAATAATAGTCTCAAAATCATTTTGTATTATATATGATA 77965

235 hr.....PhePheAsnGlnValLeuMetLeuGlyLysThr 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77966 CTACTGTGAGTTAAATTTCTAAAGAAAGCTTATTAAAGAAATGAAGACT 78015

247 SerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTy 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78016 ACCATC.....ATGAATCATAGAACCATTTGGACTAGTCTGTTCCTTTT 78059

263 rAsnIleAsp.....SerHisLeuLeuLeuSerValLeuProGlnL 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78060 TTCACACACAACTTCGTAATCATATACATCAACTTCAAAACATTTGGCTAAA 78109

277 euGluPheLysLeuLysSerAsnAspAsn...GluGluArgLeuGlnVal 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78110 ATTACTATTACTTGTAGTAATAGTAATGGAATATCATATCAACTC 78159

293 ValLysLeu.....LeuAlaLysMetPheGlyAlaLysAspSerG1 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78160 CTTAAATGAAGATAGTTTTCCTCCCGATGCAAGGTGGA...GATAGTGA 78206

306 u 306
78207 G 78207

seq_name: gb_htg8:AC022862

seq_documentation_block:
LOCUS AC022862 160962 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-343C9 map 4, WORKING DRAFT
SEQUENCE, 38 unordered pieces.
ACCESSION AC022862
VERSION AC022862.3 GI:7249278
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160962)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-343C9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160962)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
```



Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,  
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,  
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,  
 Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

COMMENT

Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L10823

Center clone name: 2504\_H2

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 135676 bases at least Q40  
 Consensus quality: 149264 bases at least Q30  
 Consensus quality: 155372 bases at least Q20  
 Insert size: 219000; agarose-fp  
 Insert size: 159671; sum-of-contigs  
 Quality coverage: 2.3 in Q20 bases; agarose-fp  
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 81 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      162: contig of 162 bp in length
*      163 262: gap of 100 bp
*      263      674: contig of 412 bp in length
*      675 774: gap of 100 bp
*      775      1187: contig of 413 bp in length
*      1188 1287: gap of 100 bp
*      1288      1965: contig of 678 bp in length
*      1966 2065: gap of 100 bp
*      2066      2205: contig of 140 bp in length
*      2206 2305: gap of 100 bp
*      2306      2596: contig of 291 bp in length
*      2597 2696: gap of 100 bp
*      2697      3096: contig of 400 bp in length
*      3097 3196: gap of 100 bp
*      3197      3778: contig of 582 bp in length
*      3779 3878: gap of 100 bp
*      3879      4519: contig of 641 bp in length
*      4520 4619: gap of 100 bp
*      4620      5136: contig of 517 bp in length
*      5137 5236: gap of 100 bp
*      5237      5877: contig of 641 bp in length
*      5878 5977: gap of 100 bp
*      5978      6662: contig of 685 bp in length
*      6663 6762: gap of 100 bp
*      6763      7437: contig of 675 bp in length
*      7438 7537: gap of 100 bp
*      7538      8181: contig of 644 bp in length
*      8182 8281: gap of 100 bp
*      8282      8720: contig of 439 bp in length

```

```

*      8721 8820: gap of 100 bp
*      8821      9297: contig of 477 bp in length
*      9298      9397: gap of 100 bp
*      9398      10235: contig of 838 bp in length
*      10236 10335: gap of 100 bp
*      10336      10985: contig of 650 bp in length
*      10986 11085: gap of 100 bp
*      11086      12119: contig of 1034 bp in length
*      12120 12219: gap of 100 bp
*      12220      13370: contig of 1151 bp in length
*      13371 13470: gap of 100 bp
*      13471      14862: contig of 1392 bp in length
*      14863 14962: gap of 100 bp
*      14963      15547: contig of 585 bp in length
*      15548 15647: gap of 100 bp
*      15648      16772: contig of 1125 bp in length
*      16773 16872: gap of 100 bp
*      16873      18277: contig of 1405 bp in length
*      18278 18377: gap of 100 bp
*      18378      19739: contig of 1362 bp in length
*      19740 19839: gap of 100 bp
*      19840      21158: contig of 1319 bp in length
*      21159 21258: gap of 100 bp
*      21259      22237: contig of 979 bp in length
*      22238 22337: gap of 100 bp
*      22338      23288: contig of 951 bp in length
*      23289 23388: gap of 100 bp
*      23389      24797: contig of 1409 bp in length
*      24798 24897: gap of 100 bp
*      24898      25704: contig of 807 bp in length
*      25705 25804: gap of 100 bp
*      25805      27367: contig of 1563 bp in length
*      27368 27467: gap of 100 bp
*      27468      28651: contig of 1184 bp in length
*      28652 28751: gap of 100 bp
*      28752      30113: contig of 1362 bp in length
*      30114 30213: gap of 100 bp
*      30214      31445: contig of 1232 bp in length
*      31446 31545: gap of 100 bp
*      31546      32766: contig of 1221 bp in length
*      32767 32866: gap of 100 bp
*      32867      34495: contig of 1629 bp in length
*      34496 34595: gap of 100 bp
*      34596      36032: contig of 1437 bp in length
*      36033 36132: gap of 100 bp
*      36133      37465: contig of 1333 bp in length
*      37466 37565: gap of 100 bp
*      37566      38460: contig of 895 bp in length
*      38461 38560: gap of 100 bp
*      38561      39981: contig of 1421 bp in length
*      39982 40081: gap of 100 bp
*      40082      41219: contig of 1138 bp in length
*      41220 41319: gap of 100 bp
*      41320      43417: contig of 2098 bp in length
*      43418 43517: gap of 100 bp
*      43518      44913: contig of 1396 bp in length
*      44914 45013: gap of 100 bp
*      45014      46461: contig of 1448 bp in length
*      46462 46561: gap of 100 bp
*      46562      47955: contig of 1394 bp in length
*      47956 48055: gap of 100 bp
*      48056      49780: contig of 1725 bp in length
*      49781 49880: gap of 100 bp
*      49881      51344: contig of 1464 bp in length
*      51345 51444: gap of 100 bp
*      51445      53900: contig of 2456 bp in length
*      53901 54000: gap of 100 bp
*      54001      55543: contig of 1543 bp in length
*      55544 55643: gap of 100 bp
*      55644      57360: contig of 1717 bp in length
*      57361 57460: gap of 100 bp
*      57461      59760: contig of 2300 bp in length
*      59761 59860: gap of 100 bp

```



63063 AAGCCAGATAGCTATGTTCCAGGATAAATACGATTAAATGATACTTT 63112

NOT RECORDED (K); NOT RECORDED (J);

NOT RECORDED (K); NOT RECORDED (J);







```

misc_feature      15116..16652
/note="assembly_fragment"
misc_feature      16753..19184
/note="assembly_fragment"
misc_feature      19285..21169
/note="assembly_fragment"
misc_feature      21270..24243
/note="assembly_fragment"
misc_feature      24344..27226
/note="assembly_fragment"
misc_feature      27327..30105
/note="assembly_fragment"
misc_feature      30206..32466
/note="assembly_fragment"
misc_feature      32567..35504
/note="assembly_fragment"
misc_feature      35605..38079
/note="assembly_fragment"
misc_feature      38180..41205
/note="assembly_fragment"

alignment_scores:
  Quality: 353.50      Length: 302
  Ratio: 3.243        Gaps: 3
  Percent Similarity: 36.093      Percent Identity: 29.139

alignment_block:
US-09-512-581-2 x AC023006 ..

Align seg 1/1 to: AC023006 from: 1 to: 162775

913 AspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysG1 929
|||||
143786 GATGAGTGTACCAAGTAGGACGATATTTGCTCAGAGCTGCATAAGC 143835
|||||

929 YLeuSerArgLeuArgLeuProLeuGluTyrMetAlaileCysAlaLeuC 946
|||||
143836 ACTTGTGAAGTACTGCTCCATCGGAGTATATGCGCATCTTTGCCCTGT 143885
|||||

946 ySAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
|||||
143886 GTGCCAAGATCCTGTGAAGGAGAGACACACGACGACATGTTTA 143935
|||||

963 ValLysAsnIleAsnValArgGluTyrLeuLysGlnHisAlaAla.V 979
|||||
143936 CTGAAAATATACGATATACGACGAGGAATACATTAAGCAGATCTATGCC 143985
|||||

979 al..... 979
:
143986 TACTGCTGAAGTAACTCAGAGTTCTCCGTGTGCAAAATATATTTCTTAATG 144035
|||||

979 ..... 979

144036 TATCCATTTTATAGTGTATTTTCTTGTGTCATGTGTTTCTATTATAGTTT 144085
|||||

979 ..... 979

144086 ATATAAACTGGGAGTTTATAGAATGCTGTCATCTAGTTATGATGTA 144135
|||||

979 ..... 979

144136 CTTAAAACTGAACCTCCTGAGTCTCTATTTATGCTGCATTAACCTGCGCAA 144185
|||||

979 ..... 979

144186 ACTGCATTTTCTTATGGGATCAGGAGTGTAGTTTAATTTGAACATATTG 144235
|||||

979 ..... 979

144236 CTCGGATGTCAGACTAAATAATTTTTCCTAGCTTTCGACGGGACCTGTT 144285
|||||

979 ..... 979

```

```

144286 GTTTAATGGGTTTAATATCTATGATACCTTTTGACITTCAGTATCTTTCAAAG 144335
979 ..... 979

144336 TGAATACAACTGCATACCAAGGTACCAGAAAAAGAAATTTGAAAGTGAAG 144385
979 ..... 979

144386 TGTAAAGTATTTTAGTATCCAGGTATCATGTTGTGCCCTAATTTAGAATA 144435
979 ..... 979

144436 ACTGAGGAAAAAATGAAACTTAAGTGGTTAGAACTCAGTCCACTAAGG 144485
980 .....SerGluLysLeuLe 984
144486 CATTTTATAAAGGTGTTTTTAATAATGTTTTTACTTTCAGAGAAATATT 144535
984 uSerLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
|||||
144536 ATCAGTGTGCTGCAATATGATGTTCCATACATGATTCACCTGCTAGCCC 144585
|||||

1001 isAspProAspTyrValLysValcInAspIleGluInLeuLysAspVal 1017
|||||
144586 ATGATCCAGATTTTACAAGATCAAGATGTTGATCAGCTTCGTGATATC 144635
|||||

1018 LysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGluAs 1034
|||||
144636 AAAGAGTAAGTC.....TTGTTGTTTTTCAAGAACTCTAA 144670
1034 nAsn 1035
|||||
144671 AAAT 144674

seq_name: gb_htg14:AC068352
seq_documentation_block:
LOCUS AC068352 145450 bp DNA HTG 25-JUN-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-168F19 map 4, WORKING DRAFT
SEQUENCE, 31 unordered pieces.
ACCESSION AC068352
VERSION AC068352.2 GI:8705054
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145450)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-168F19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145450)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bohuslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galligan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lechoczek,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```











198 ThrValLeu.....ValAsnLeuValProAlaHi 207  
 1475 TTAATTTTAAATAAGTTTGTGACCTATAACCCGAATGAAATTCCTGAAGG 1426  
 207 sYsAsnLeuAsnLysGln...AlaTyrAspLeuAlaLysAlaLeuLeuL 223  
 1425 TTTGAACGGTACTCTCGACTGGCTGGCTACGAAGTTAGTCTAAATTTGTGTG 1376  
 223 ysArgThrAlaGlnAlaLeuGluProTyrIlePheThrPheAsnGln 239  
 1375 ATACTTATCCATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG 1326  
 240 ValLeuMetLeu.....GlyLysThrSerIleSerAspLeu 251  
 1325 ATATTTCATGAACCACTAATGATGATGATGATGATGATGATGATGATGATG 1276  
 251 uSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerH 268  
 1275 GTGAGTGAATTAACATAACTGGTATTTAGACTTTGGGAACAGTACCCG 1226  
 268 isLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsn 284  
 1225 AACTGATCAATGACGTAATCGGATTTATATACCACGAATATCTTCGGAA 1176  
 285 AspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheG 301  
 1175 AAGGAGCTTTTCAGAAAGAAAGCTACAAAATTAATTTGGTCAAACTACTTAC 1126  
 301 yAla...LysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnC 317  
 1125 CTCATATTCGGATTTAAATTTGTTTTCACACATTCCTCCGATACGTTCAAGG 1076  
 317 ystyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333  
 1075 CATGGATTTCCAAAATGCTGATATTAGTCCCGATGCTAGAGTAGAGTGG 1026  
 334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350  
 1025 ACAGAATCAATTCCTCCAAATCATTTGCCAGGAGAGAGACATATCCAAAGA 976  
 350 pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367  
 975 ACTGAATCAGGCTGGCAAAACCTTTATGATCTGATCTCGTGATC 926  
 367 rgHisAspValIleValSerIle.....ValThrAlaAlaLys 379  
 925 GCAGGACTTCTGTATGATTTTAAATAAAGTTCCCGTTACTGAAATTTGG 876  
 380 LysAspIleLeu.....LeuValAsnAspHisLeuLeuAsnPheValar 394  
 875 AAAAATATAACCAATAAAGCCATTTATACATCTTTACTTCTATTTGGCCCG 826  
 394 gGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetG 411  
 825 AGAA.....AAACATAGAGGAGGTAGGAGACTCTGCATTAACA 788  
 411 lyLeuAlaGlnIleTyrLysLysTyrAlaLeuGln.....SerAla 424  
 787 CTATGCCCAAGTTTACTCAAATTCATTTGAATGAGATCGAAGGACATAC 738  
 425 AlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLe 441  
 737 CAAATATAAGAAATTTGGGAAATCATPAGACACATACCATCCATCTTATA 688  
 441 uHisIleTyrTrpGlnAspSerIleAspArgLeuLeuValGluArgI 458  
 687 CAATCTCTATTATCAATGATTTTAAACATCAATGACAGTCCGACGCG 638  
 458 lePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMet 474  
 637 TAATTTTCGAATACCTACTACCGTTTGAACCCAGATACCATGATAAAGAGTC 588

475 LysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAl 491  
 587 CACAGACTGCTAACGGTTTGTGCACATTTTGTGATAAGAGGCATTCACCTC 538  
 491 aLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysA 508  
 537 ATTTTTCGGCTTTCAATGCAAGCAAAATCAAAATATCTTCGGCTATATCCA 488  
 508 spLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAla 524  
 487 ATATATATTGATTTTACTAAATTTTGAACAATCAAGAAAGATGAGCTCG 438  
 525 IlePheSerLysValMetVal.....Ile 532  
 437 TCTCAAGGTCCTATTGTAATGAACAAGTACACCAAACTCTCCAGTGGTT 388  
 532 eThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysP 549  
 387 AGCTTCTGGCCTTCTGACTCTACGAAGCAATCGACGCCCTCGAAACAA 338  
 549 heThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluVal 565  
 337 TAAACAATTT.....AACGAGCAAGAAATCTTTTATTATTAAATGCC 294  
 566 LeuValSerProThrCysSerCysLysGlnAlaGluGluCysValArgG 582  
 293 TGTGTCCACCAACGATATTCGGTTTCTTACTTTTAAAAATTTGTTACAATGA 244  
 582 uIleThrLysLysLeuGlyAsnProLys..... 591  
 243 ATTGGTCAGTAAGTTTACAAACGCCAGGACTATTTTAAAAATATAATATAT 194  
 592 .....GlnProThrAsnProPheLeuGluMetIleLys 602  
 193 CTACTGGTCTTCCATTTATGCCCAGGTGAT...ATCCCCAAGGTAATACAA 147  
 603 PheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSe 619  
 146 ATTTTCTCTTTTAGAGCCTCACCATAATTTATATGTTATCCCAACATCAG 97  
 619 rAlaLeuIleLysGlnValAsnLysSer 628  
 96 TGTGTTTATGACCTATCTAATAATTTCT 69

seq\_name: gb\_pr5:AK023592

seq\_documentation\_block: 1489 bp mRNA PRI 29-SEP-2000  
 LOCUS AK023592 Homo sapiens cDNA FLJ13530 fis, clone PLACE1006248, highly similar  
 DEFINITION to Homo sapiens mRNA for KIAA0648 protein.

ACCESSION AK023592  
 VERSION AK023592.1 GI:10435570  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib:PLACE1  
 clone:PLACE1006248.

# ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE

1 (sites)  
 Isoqai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
 Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,  
 Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y.,  
 Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K.,  
 Masuho, Y. and Kanehori, K.  
 NEDO human cDNA sequencing project

# TITLE

Unpublished (2000)

# JOURNAL

2 (bases 1 to 1489)

# REFERENCE

Direct Submission

# AUTHORS

Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao

Isoqai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,

Tel:81-438-52-3951, Fax:81-438-52-3952).



```

E"
misc_feature
/feature="CHL12"
/feature="PS00017 ATP/GTP-binding site motif A"
join(7008..7016,7173..8078)
/feature="SEC14"
/feature="SEC14"
join(7008..7016,7173..8078)
/feature="SEC14"
/feature="YMG582.04, SEC14 gene (PIT1), len: 304, CAI: 0.31,
SC14_YEAST P24280 SEC14 cytosolic factor
(phosphatidylinositol /phosphatidylinositoltransfer
protein)"
/codon_start=1
/product="Sec14p"
/protein_id="CAA89225.1"
/db_xref="GI:807961"
/translation="MVTOQKEFLSYQNCPPDLPCTPGNLSAQEKALAEKRLK
EDAGTIERDDSTLRLFRARKEFDVLAKEMFENCEKWRKDYGTDTILQDFHYDEKPL
IAKPYOYVHKDKGRPVYFEELGAVNLHMKVTSERMLKNLWVEYVVOYRLP
ACSRAGHLVETSCITMDLKGISISAVSMYSVREASVISONYYPERMGKEYINAP
FGFSFAFLKFLPDVTVSKIFILGSSYQKELLKOIPAENLPVAFGGKSEVDESKGG
LYLSDIGMWRDPKYIGPEGEAPEAFSMK"
7017..7022
/feature="splice donor sequence, gtatgt"
/feature="7128..7134
/feature="splice branch sequence, tactaac"
7170..7172
/feature="splice acceptor sequence, tag"
complement(8731..11646)
/feature="NAM7"
complement(8731..11646)
/feature="NAM7"
/feature="YMG582.05c, NAM7 gene (UPF1), len: 971, CAI: 0.21;
NAM7_YEAST P30771 NAM7 probable helicase; PS00017
ATP/GTP-binding site motif A"
/codon_start=1
/product="Nam7p"
/protein_id="CAA89226.1"
/db_xref="GI:807962"
/translation="MVGSGHTPYDISNSPDVNVQPAQLNSTLVEDDDVDNOLFEE
AQVTFGRSPASNSCAICGIDSAKVICNSCKKWCNTKNGTSSSHIVNHLVLS
HNVSVLHSPDSGLDTLCYNGRKNVLLGFVSAKSAVVVLLCRIPCAOTKANW
DTDQOPLLEDQLLSWAEOPTREKILKRLITPSQISKLEAKWRSNKDATTINDIDA
PEQEARPELLRLYODAYEQSYGLPKLEADYDKQLKESQALEHISVDSWALNRR
HLASTFTFSENEUKVAIGDEMILVYSCMOHPDWEGRYIVRLPNSFOTFTLELKP
SKTPPTHTTGTAEFTWKGTSYDRMODALKFAIDKKSISGYLYYKILGHVVDLIS
FDVPLPKFESIPFNAQLNSSQSNVSHVLRPLSLITQPGTGKTVYATIVYHLSKI
HKDRLVLCAPSNAVVDHLAALDKLKVRLTAKSREDVSSVNLALHNLVGRGAK
GELNLLKLDKDELSADTKRFVYKLVKTEAILNKADVCCCTCVGAGDXRLDTKP
RVLIDESTQASEPECLIPVIGAKQVILVGDHQQGLPVILLERKAADAGLKOSLPERL
ISLGHVPIREVOYRNNPYSLEFPNMEYEGSLONGVTIEQRTVPSKPPPIRGIPM
MFVANGREISANTSTFLNRTEAMNERIITKFRDGVKPEQIGVITPYEGORAYIL
QYOMNGSLDKDLYIKVEVASDAFOGREKDYIILSCVRANEQOAGIFURDPRRLNVG
LTRAKYGLVILGNLANLTLNHLHIFREKGLVCEGLDNLQCLTQVLRPOPQRK
TERPNAQNVSEMGDFPKFQDFDAQSVSVSGQIGDFGNFVNTLSSVINNEYW
NFENFASFQONKRNEDIRNLYQIEASHLSNFARELQREQKHELSKDFSLNIG"
/feature="NAM7"
complement(10336..10359)
/feature="PS00017 ATP/GTP-binding site motif A"
complement(12098..13114)
/feature="MBR3"
complement(12098..13114)
/feature="MBR3"
/feature="YMG582.06c, MBR3 gene (ISF1)len: 338, CAI: 0.14;
MBR3_YEAST P32488, MBR3 protein (increasing suppression
factor 1) (isf1 protein)"
/codon_start=1
/product="Mbr3p"
/protein_id="CAA89227.1"
/db_xref="GI:807963"
/db_xref="SWISS-PROT:P32488"

```

```

/translation="MIASEIFERGVQDFQCDQDEYEDTQVQSLGSLNDLNFVSKL
ASFSQNSKSNHSHSSNTNAGNTIGHHNRKIKTAPHHLYGFVPANSTN
NSNEPIRSPRIIRANSSTLIHQLSROSTROSSLDGADSCDFHKCIKIPRSRHSCTG
IPTHLYGLEKYVSSSELDLAVANDQSNLDLTSPITSVSTPASNSYLNLSAAAYPS
SYLTNRKNKRLKLI SHGKISSNNVPGHGNLHNYHRETPSNLRRSFSLLSNGSSS
PQTRNNYSYNSLVSPNSPNSLNTSVASSNEESPHTSNCLEERNPRKRSFKLSLAS
SPSN"
complement(13788..14144)
/feature="YMG582.07c, unknown, len: 118, CAI: 0.12"
/codon_start=1
/product="unknown"
/protein_id="CAA89228.1"
/db_xref="GI:807964"
/db_xref="SWISS-PROT:O04276"
/translation="MTIAETIRQSFYPLFGNVLGHWPVPRVPRKEIKMLLFTAN"

alignment_scores:
Quality: 377.00 Length: 676
Ratio: 0.990 Gaps: 18
Percent Similarity: 56.361 Percent Identity: 21.302

alignment_block:
US-09-512-581-2 x SC9582X/rev ..
Align seg 1/1 to reverse of: SC9582X from: 1 to: 38661

12 LyslethrThrProGlyValLysGluIleSerAspLys....IleSe 27
||||| : : : : : ||||| : : : : : |||||
207 AAACCTGAAGTTTAACTCACTTAAATATATCTACCTCCGATGAGTATC 1978
27 rLysGluMetValArgArgLeuLysMetValValLysThrPheMetA 44
: : : : : ||||| : : : : : : : : :
1977 AACAAATGAAGTATTAGATCGTTTAAAGCTTTACATGAGGAATAGCCT 1928
44 sMetAspGlnAspSerGluGluLysGluLeu.....TyrLeuAsn 58
: : : : : ||||| : : : : : |||||
1927 CTCTGGATCAAGACATACAGATTTAACGGGATTAGATAATATAGGAC 1878
59 LeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAs 75
||||| : : : : : |||||
1877 GCA.....CTAGTAGCAGAAAATTTAAACATAAAGACGTGGG 1837
75 pValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyr 92
: : : : : ||||| : : : : : |||||
1836 GATCCGGCTTTCACCGCTTGTCTTAAGGATATCTGAGATATATAG 1787
92 laProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMet 108
||||| : : : : : ||||| : : : : : |||||
1786 CTCAGATGCGCTTATACGGATGCT...CAACTAACGGACATTTTCAAG 1740
109 PheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPh 125
: : : : : ||||| : : : : : |||||
1739 CTGTGCTCTCACAAATTTGAACAATTAGTGACCAAGAAACGGCTATCA 1690
125 eAsnArgPheTyrLeuLeuGluAsnIleAlaTyrValLysSerTyrA 142
: : : : : ||||| : : : : : |||||
1689 CATCCACAACACTTATTGTATCATACTAGCTAGATAAGATCAATCG 1640
142 snIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyr 158
: : : : : ||||| : : : : : |||||
1639 TATTATTAGCGGATTACCTAGTTCAACAATTTTATTCGAACTATTC 1590
159 Arg.....ThrLeuPheSerVa 164
: : : : : ||||| : : : : : |||||
1589 CATATTTTTTACGATCTTAAACAGAGCTTCCCGCAAGGCTCTTCAATGT 1540
164 lIleAsnAsnGlyHisAsnGlnLysValHisMetValAspLeuM 181
||||| : : : : :
1539 AATAGCGGT.....ATTC 1526
181 etSerSerIleCysGluGlyAspThrValSerGlnGluLeuLeuAsp 197
: : : : : ||||| : : : : : |||||
1525 TCGGTGAAGTCATCTCTGATTCGATTCGTTCCCTAGAAAGTACTAAGA 1476

```







```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA448113
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159306 bases at least Q40
Consensus quality: 161143 bases at least Q30
Consensus quality: 162084 bases at least Q20
Insert size: 162842; sum-of-contigs
Insert size: 163676; 2.5% error; agarose-fp
Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality
coverage: 4.42x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 20261: contig of 20261 bp in length
* 20262 20361: gap of 100 bp
* 20362 38140: contig of 17779 bp in length
* 38141 38240: gap of 100 bp
* 38241 85018: contig of 46778 bp in length
* 85019 85118: gap of 100 bp
* 85119 93385: contig of 8267 bp in length
* 93386 93485: gap of 100 bp
* 93486 96470: contig of 2985 bp in length
* 96471 96570: gap of 100 bp
* 96571 137301: contig of 40731 bp in length
* 137302 137401: gap of 100 bp
* 137402 147035: contig of 9634 bp in length
* 147036 147135: gap of 100 bp
* 147136 149710: contig of 2575 bp in length
* 149711 149810: gap of 100 bp
* 149811 163642: contig of 13832 bp in length.
*
* Location/Qualifiers
* 1..163642
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="13"
*   /clone="RP11-448I13"
*   /clone_lib="RPCI-11.2"
* 1..20261
*   /note="assembly_fragment:00330"
*   clone_end:T7
*   vector_side:left
* 20362..38140
*   /note="assembly_fragment:00335"
*   fragment_chain:1
* 38241..85018
*   /note="assembly_fragment:01397"
*   fragment_chain:1
* 85119..93385
*   /note="assembly_fragment:01083"
*   fragment_chain:1
* 93486..96470
*   /note="assembly_fragment:00197"
*   fragment_chain:1
* 96571..137301
*   /note="assembly_fragment:00284"
*   fragment_chain:2
* 137402..147035
*   /note="assembly_fragment:00874"
*   fragment_chain:2

```

```

misc_feature 147136..149710
              /note="assembly_fragment:00391"
              fragment_chain:3"
misc_feature 149811..163642
              /note="assembly_fragment:00398"
              fragment_chain:3"
BASE COUNT 54756 a 31368 c 29374 g 47337 t 807 others
ORIGIN
alignment_scores:
  Quality: 390.50      Length: 349
  Ratio: 3.719        Gaps: 3
  Percent Similarity: 30.086  Percent Identity: 29.799
alignment_block:
  US-09-512-581-2 x AL353724/rev ..
  Align seg 1/1 to reverse of: AL353724 from: 1 to: 163642
104 LysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120
|||||
26467 AAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATAC 26418
|||||
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu..... 133
|||||
26417 AAAGGCCACAAATCAATAGTATTTTATTACTTGAGGTAAGCAATA 26368
|||||
133 ..... 133
26367 TCTGTATCTTGAGATGACATTTTAAACTGATTTTGTGTTGTATATT 26318
|||||
134 .....As 134
26317 TTAATAATCAAAATGCTTAACTTTTACCTTTTACACCTTATTTTAGAA 26268
|||||
134 nileAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerA 151
|||||
26267 CATGCTGGTCAAGCATATATACATATGCTTTCAGTTAGAGATAGCA 26218
|||||
151 snGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIle..Asn.. 166
|||||
26217 ATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAGTAA 26168
|||||
166 ..... 166
26167 GTTTATTTTAAAGTATGTAAACATTTAAAAAAGTAACAAATTTTGAAC 26118
|||||
166 ..... 166
26117 TTGTAGGTGAATCTGCATAAAGCTTTACAATGAATATTAGAGGCTGAGA 26068
|||||
166 ..... 166
26067 GATGTAACCTTTTAACACACACTTTTATTTTATACATGTTTCAAAAATGAATA 26018
|||||
166 ..... 166
26017 GAGCTTAATTTTGTATTATATATGTAATTATTGTTAGAGGAGTGCCATAAG 25968
|||||
166 ..... 166
25967 GCCAGTGTAAATTTTTCGATTCGATTTTTCCTATAAAACATACAT 25918
|||||
166 ..... 166
25917 ACTGACTTCTTAATGCATATGGCTATTTTGAGCTTTTCTGTATGCTTTAA 25868
|||||
166 ..... 166
25867 AATTATTTACTGGAATTAACCTCAGAGCAATTCATTTTCTGAATTTAAG 25818
|||||
166 ..... 166

```

FEATURES  
Source



```
repeat_region /note="Alu repeat: matches 1. .308 of consensus"
75593. .75883
repeat_region /partial
/note="Alu repeat: matches 308. .1 of consensus"
75974. .76251
repeat_region /partial
/note="Alu repeat: matches 294. .1 of consensus"
76551. .76608
repeat_region /note="MIR element fragment"
78482. .78832
repeat_region /note="MER11B element fragment"
78574. .78743
repeat_region /note="MER11A element fragment"
78941. .79021
repeat_region /note="MER11B element fragment"
79249. .79709
repeat_region /note="MER11A element fragment"
80233. .80302
repeat_region /note="SVA element fragment"
80355. .80535
repeat_region /partial
/note="Alu repeat: matches 213. .20 of consensus"
80563. .80619
repeat_region /partial
/note="Alu repeat: matches 277. .221 of consensus"
80651. .82294
repeat_region /note="SVA element fragment"
82753. .83043
repeat_region /partial
/note="Alu repeat: matches 308. .1 of consensus"
83149. .83296

alignment_scores:
Quality: 390.50 Length: 349
Ratio: 3.719 Gaps: 3
Percent Similarity: 30.086 Percent Identity: 29.799

alignment_block:
US-09-512-581-2 x HS267P19 ..
Align seg 1/1 to: HS267P19 from: 1 to: 113704
104 LysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120
|||||
22830 AAGGATATATTTGCTTTATACAGACAGAGTTCAAGGGCTAGAGGATAC 22879
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu..... 133
22880 AAAGAGCCCAACATTCATAGATATTTTATTACTTGAGGTAAGCAATA 22929
133 ..... 133
22930 TCTTGATCTTGAGATGACATTTTAACTGATTTTCTGTTGATATATT 22979
134 .....As 134
22980 TTAATAATCAAAATGCTTAACCTTTCACCTTTTACACCTTATTTTAGAA 23029
134 ntleAlatrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerA 151
|||||
23030 CATTCCTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGCA 23079
151 snGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIle.Asn.. 166
|||||
23080 ATGAATTTTCCACCCAGCTATACAGAACCTTATTTTCAGTTATTAAGTAA 23129
166 ..... 166
23130 GTTATTTTATTAAGTATGATACATTTAAAAAAGGTAACAAATTTTGAAC 23179
166 ..... 166
23180 TWTAGTGAACTGCATAAACTGTTACAAATGATATTAAGAGGCTGAGA 23229
```

```
166 ..... 166
23230 GATGTAACCTTTTAAACACACTTTATTTTATACATGTTTCAAAAATGAATA 23279
166 ..... 166
23280 GAGCTTAATTTTGTATATATATATTAATATTGTTAGAGAGTGGCATAAG 23329
166 ..... 166
23330 GCCAGTGTAAATTTTGGATTGCATTTATTTGCTATATAAACATACAT 23379
166 ..... 166
23380 ACTGACTTCTTAATGCATATGGCTATTGAGCTTTTGTGTATGCTTTAA 23429
166 ..... 166
23430 AATTATTACTGGAATTAACCTTCAGAGCAATTCATTTTCTGAATTTAAG 23479
166 ..... 166
23480 GCAGAACTCAAGTTGCTTTTGATGTAATTTACTTTTGTGAGAGCTGGT 23529
166 ..... 166
23530 TAAAAATTAGGAAAAAATCTAAGAAATGTAGTAAAAAATATAATTTGTA 23579
166 ..... 166
23580 TACTTTTGTATGATGCGGAGAGAGGGAAGCTTTAAATGAATATACATAG 23629
166 ..... 166
23630 AATTTTATTAATATTTTACTATTTTGTGTAAGCAATAAGCTTTGTCA 23679
166 ..... 166
23680 GTGTCATCTTAAGAGTAATCTCTGATATATCTCAGTTGTAAATGAAACAAA 23729
167 .....AsnGlyHisAsnGlnLysValHisMetHisM 177
|||||
23730 ATCTGTTTGAATTCAGCAATGGCCACATCAGAAAGTCCATATGCACA 23779
177 evValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGln 193
|||||
23780 TGTAGACCTTATGAGCTCTATTTTGTGAAGGTGATACAGTGTCTCAG 23829
194 GluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLys 208
|||||
23830 GAGCTTTGGATACGGTTTGTAGTAAATCTGGTACCTGCTCATAAG 23874
```

seq\_name: gb\_htg20:AL353724

```
seq_documentation_block:
LOCUS AL353724 163642 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-448113, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL353724
VERSION AL353724.3 GI:9930876
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163642)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Aug 27, 2000 this sequence version replaced gi:9926534.
```



such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. 'Alu'), where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 267P19 is at 1 in this sequence. The true right end of clone 26H23 is at 24539.

The true left end of clone 45710 is at 113606.

267P19 is from the human PAC library.

## FEATURES

source

source	1. .113704 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="13" /map="13q12-13" /clone="XX-267P19" 1. .235 /partial /note="Alu repeat: matches 252. .1 of consensus" 370. .662 /partial /note="Alu repeat: matches 308. .1 of consensus" 4849. .5140 /partial /note="Alu repeat: matches 308. .1 of consensus" 5367. .5650 /note="Alu repeat: matches 1. .308 of consensus" 6080. .6187 /note="2 copies of 54 mer 87 % conserved" 8133. .8401 /note="MER7A element fragment" 8623. .8748 /note="L1 element fragment" 9617. .9711 /partial /note="Alu repeat: matches 305. .210 of consensus" 9724. .9876 /partial /note="Alu repeat: matches 158. .1 of consensus" 10567. .10594 /note="14 copies of 2 mer 93 % conserved" 10616. .10900 /partial /note="Alu repeat: matches 300. .1 of consensus" 10960. .10985 /note="13 copies of 2 mer 92 % conserved" 11912. .12100 /partial /note="Alu repeat: matches 268. .65 of consensus" 17030. .17359 /note="Alu repeat: matches 1. .306 of consensus" 17639. .17888 /partial /note="Alu repeat: matches 34. .306 of consensus" 19151. .19980 /note="L1 element fragment" 19981. .20264 /partial /note="Alu repeat: matches 308. .1 of consensus" 20298. .21190 /note="L1 element fragment" 21199. .21488 /partial /note="Alu repeat: matches 308. .1 of consensus" 21538. .22226 /note="L1 element fragment" 25636. .25989 /note="Alu repeat: matches 1. .308 of consensus" 26902. .27167 /note="2 copies of 133 mer 93 % conserved" 27850. .27930 /note="L1 element fragment" 28002. .28043 /note="L1 element fragment"
--------	---

repeat_region	28323..28780 /note="L1 element fragment"
repeat_region	29176..29449 /note="Alu repeat: matches 1..306 of consensus"
repeat_region	29904..30192 /note="Alu repeat: matches 1..308 of consensus"
repeat_region	30354..30663 /note="L1 element fragment"
repeat_region	30683..30972 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	30977..31204 /note="L1 element fragment"
repeat_region	34819..35121 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	37059..37282 /note="MLT2B2 element fragment"
repeat_region	37453..37506 /note="MLT2B2 element fragment"
repeat_region	38414..38706 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	39516..39563 /note="24 copies of 2 mer 81 % conserved"
repeat_region	41244..41535 /note="Alu repeat: matches 1..308 of consensus"
repeat_region	43994..44242 /partial
misc_feature	/note="Alu repeat: matches 44..308 of consensus"
repeat_region	complement(44369..44677) /note="match: Z17151 DNA segment containing (CA) repeat"
repeat_region	complement(44479..44510) /note="16 copies of CA 100% conserved; Polymorphic by comparison with Z17151"
repeat_region	44702..44889 /note="2 copies of 94 mer 85 % conserved"
repeat_region	45935..46229 /note="Alu repeat: matches 1..308 of consensus"
repeat_region	46910..46988 /note="L1 element fragment"
repeat_region	47081..47119 /note="L1 element fragment"
repeat_region	47168..47431 /note="L1 element fragment"
repeat_region	47615..47902 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	47922..48213 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	49426..49717 /note="Alu repeat: matches 1..308 of consensus"
repeat_region	50248..50533 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	51097..51380 /note="Alu repeat: matches 1..304 of consensus"
repeat_region	62318..62543 /partial
repeat_region	/note="Alu repeat: matches 302..66 of consensus"
repeat_region	63358..63613 /partial
repeat_region	/note="Alu repeat: matches 306..34 of consensus"
repeat_region	66191..66236 /note="23 copies of 2 mer 80 % conserved"
repeat_region	68771..69062 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"
repeat_region	70282..70559 /partial
repeat_region	/note="Alu repeat: matches 308..1 of consensus"



```

misc_feature 1..2333
/note="assembly_fragment:02826
fragment_chain:1"
2434..4472
/misc_feature /note="assembly_fragment:00159
fragment_chain:1"
4573..6623
/misc_feature /note="assembly_fragment:00292"
6724..8757
/misc_feature /note="assembly_fragment:02097"
8858..11498
/misc_feature /note="assembly_fragment:02560"
11599..13673
/misc_feature /note="assembly_fragment:03440"
13774..16148
/misc_feature /note="assembly_fragment:03977"
16249..195380
/misc_feature /note="assembly_fragment:04183"
16249..195380
BASE COUNT 55503 a 40458 c 40664 g 58048 t 707 others
ORIGIN

alignment_scores:
Quality: 401.50 Length: 361
Ratio: 3.788 Gaps: 2
Percent Similarity: 29.363 Percent Identity: 29.086

alignment_block:
US-09-512-581-2 x AL358892 ..
Align seg 1/1 to: AL358892 from: 1 to: 195380

103 LeuLysAspPheMetPheLeuArgGlnLeuLysGlyLeuGluAs 119
|||||
170504 TTAAGGATATATTTATGTTTATAACAAGGCACTAAAGGGACTAGAACA 170553

119 pThrLysSerProGlnPreAsnArgTyPheTyrlLeuLeuGlu..... 133
|||||
170554 TACAAGAGCCCTCAATTTAAATAGTATTTTATTACTTGAGGTAAGCA 170603

133 ..... 133

170604 ATACATTATTACTTGAGTTGATTCTTTTCTTCTATATATCTTTTACC 170653

134 .....AsnIleAlaTr 137
|||||
170654 AACCAATATATTACTTTCCCTACTTCTTTCTTTTGAACATTTGCATG 170703

137 pValLysSerTyAsnIleCysPheGluLeuGluAspSerAsnGluIleP 154
|||||
170704 GGTAAATCATATACATATGCTTTGAGTTAGAAGATAGCAATGAATCT 170753

154 heThrGlnLeuTyArgThrLeuPheSerValIle..... 165
|||||
170754 TTACTCAACTTTACAGAACATTATCTCAGTTATAAAGTAAAGTTGATTT 170803

165 ..... 165

170804 TAAATAACATTTCTTTTAAATGTAAGCTTTTCTGTTTACTATGAATCC 170853

165 ..... 165

170854 ACATAAACTGTTATAAATATTGCTTAGGATTATAAATAACTTACAAT 170903

165 ..... 165

170904 TATTCTCTATATAGTTCTAAATGAATAGATTATATATATATTTGTT 170953

165 ..... 165

170954 AGAAGAGCAAAATAGGCTAAGTTTGGATTATATCTTCTCTGATTACAT 171003

165 ..... 165

```

```

171004 ACTATGTTAACTTTTGTGTGTCATTGAGATGTTTCCCATGAGCTCTGG 171053
165 ..... 165
171054 CTGTGCTAGAACTCTCTGCAAAACCAGGCTGGCCTCAGACTCAGAGCCT 171103
165 ..... 165
171104 GCTTCTGCTTCCCACTGCTGAGATCAATGGCATGGCATTGACCATCATCTGA 171153
165 ..... 165
171154 CTTATTTGACTTTTAAATATATATCATGTTGTTGGAAAATTTTGTATTATC 171203
165 ..... 165
171204 TATAAAATGTAATAGAAATTTTAAATTTAAGAACTAAGATTGCATTTC 171253
165 ..... 165
171254 TTTAATTACTTTGGTGAAGAAATGCTAAATATTAGGGGAAAGTAATTTA 171303
165 ..... 165
171304 GTAATATAGTTTAAATAATACATTTTATTATTGTAAGACCAGAGAGAAA 171353
165 ..... 165
171354 GAAGCTTGACAGTAGTACAGATTATCTCACCACITTTTCATAAAGTATAT 171403
165 ..... 165
171404 GTTACTACTACCATATCTATCTCAGTTGAAATAAAATAAAGTTTGGG 171453
166 .....AsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuM 181
|||||
171454 TTACAGCAATGGCCCAATCAAGAAAGTTTCATATGCACATGGTGGACCTCA 171503
181 etSerSerIleCysGluGlyAspThrValSerGlnGluLeuLeuAsp 197
|||||
171504 TGAGCTCATCATTTTGTGAAGTGATACCGTATCTCAGGAGCTCTTAGAT 171553
198 ThrValLeuValAsnLeuValProAlaHisLys 208
|||||
171554 ACAGTTTGTAGTAAATCTGTTACCTGCCATAAG 171586
seq_name: gb_pr8:HS267P19
seq_documentation_block:
LOCUS HS267P19 113704 bp DNA 22-NOV-1999
DEFINITION Human DNA sequence from cosmid 267P19, BRCA2 gene region chromosome
13q12-13 contains polymorphic CA repeat.
ACCESSION 275889
VERSION 275889.1 GI:1430780
KEYWORDS 13q12-13; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113704)
Odell,C.
Direct Submission
Submitted (04-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT:
This sequence is not the entire insert of clone 267P19. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,

```



```
904 yrglnLeuCysAlaLeuAlaIleAsn..... 912
:|||||
92230 TTCAGCTCTGTCACATTGTTATTAAATAGTAACAATCTTATTTTGTG 92181
912 ..... 912
92180 AATTACTTATCTTCAAAGTTGAATACATTACTTTTCTATGGATACAAATC 92131
912 ..... 912
92130 CATTATAATGTGAATATATGACAAAAGTCACACTACTATTATTACCTGTT 92081
912 ..... 912
92080 TAACTGCTTCCAAATTATTACATTATTATTCTCCCATATAATTATCTTC 92031
912 ..... 912
92030 TTTTAAATTCAGACACATAAAATACTTAAACCTGCTGTTTATTATTAAT 91981
913 ..... Asp.GluCysTyr 916
91980 AATAGGTGTCGAATTAAGTTTCTGCTGCTGTAGGATGAGTGTTC 91931
917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLe 933
91930 CAAGTAGGAGAGATATTTGCTCAGAGCTGCATAGGCCACTTGTGAAGTT 91881
933 uArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 950
91880 ACTGCTCCCAATGGAGTATATGGCATCTTGGCTTGTGTGCCAAAGATC 91831
950 roValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
91830 CTGTGAAGGAGAGAGACACGACGACGAATGTTTACTGAAATAATC 91781
967 AsnValArgArgGluTyrLeuLysGlnHisAlaAla.Val..... 979
:|||||
91780 AGTATAGCAGGAATACATTAGCAGAACTCTATGGCTACTGGTAAGTA 91731
979 ..... 979
91730 ACTCAGAGTTCTCCGTGCAAAATATATTTTGTAAATGTATCCATTTTA 91681
979 ..... 979
91680 GGTATTTTCTGTCATGTGTTCTTATATAGTTTATAATAAACTG 91631
979 ..... 979
91630 GGAGTTTATAGAAATGCTGTCATGTAGTTATGATGTACTTAAAACTGA 91581
979 ..... 979
91580 ACTCCTGAGTCCCTATTATGTCGATTAACACTCGCAAACTGCATTTTC 91531
979 ..... 979
91530 CTATGGGATCAGGAGTGTAGTTAAATTTGAACATATGCTCGGATGTCAG 91481
979 ..... 979
91480 ACTAAATAATTTTCTTCTAGCTTTGCCAGGACCTGTTGTTTAAATGGTT 91431
979 ..... 979
91430 TAATATCTGATACCTTTGACTTCAGTATCTTTCAAAAGTGAATACAACT 91381
979 ..... 979
91380 GCATACCAGTACCAGAAAAGAAATTTGAAAGTGAAGTCTAAAGTATTT 91331
979 ..... 979
```

```
91330 TAGTATCCAGGTATCATGTTGTGCTTAATTTAGAAATAACTGAGGAAAA 91281
979 ..... 979
91280 AATGAAAACTTAAGTGGTTAGAACTCAGTCCACTAAGGCATTTTATAAAG 91231
980 ..... Ser.GluLysLeuLeuSerLeuLeuPr 988
91230 GTTGTGTTTAAATAATGTTTACTTTCAGAGAAATATTATCAGCTGTGCC 91181
988 oGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1005
91180 TGAATATGTAGTCCATACATGATTCACCTGCTAGCCCATGATCCAGATT 91131
1005 yrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeu 1021
91130 TTACAAGATCACAAAGATGTTGATCAGCTGCTGATATCAAGAGTAAGTC 91081
1022 TrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsn 1035
91080 ..... TGTGTTGTTTTCAGAACTCTAAAAAT 91054
```

seq\_name: gb\_htg21:AL358892

seq\_documentation\_block:

LOCUS AL358892 195380 bp DNA HTG 06-MAR-2001

DEFINITION Mus musculus chromosome 5 clone RP21-417G6, \*\*\* SEQUENCING IN

PROGRESS \*\*\*, in unordered pieces.

ACCESSION AL358892

VERSION AL358892.12 GI:11493289

KEYWORDS HTG; HTGS-PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

Wall, M

Direct Submission

Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequests@sanger.ac.uk

On Nov 30, 2000 this sequence version replaced gi:11414592.

----- Genome Center

Center: UK Medical Research Council

Center code: UK-MRC

Web site: http://mrcseq.har.mrc.ac.uk

Contact: mouseq@har.mrc.ac.uk

----- Project Information

Center project name: dm417G6

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 190612 bases at least Q40

Consensus quality: 192224 bases at least Q30

Consensus quality: 193166 bases at least Q20

Insert size: 194680; sum-of-contigs

Insert size: 191947; 3.7% error; agarose-fp

Quality coverage: 7.33x in Q20 bases; sum-of-contigs Quality

coverage: 8.85x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

FEATURES

Location/Qualifiers

1..195380

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="5"

/clone="RP21-417G6"

/clone\_lib="RPCI-21"



```

* 1519 2614: contig of 1096 bp in length
* 2615 2714: gap of 100 bp
* 2715 4205: contig of 1491 bp in length
* 4206 4305: gap of 100 bp
* 4306 7004: contig of 2699 bp in length
* 7005 7104: gap of 100 bp
* 7105 9239: contig of 2135 bp in length
* 9240 9339: gap of 100 bp
* 9340 11600: contig of 2261 bp in length
* 11601 11700: gap of 100 bp
* 11701 13258: contig of 1558 bp in length
* 13259 13358: gap of 100 bp
* 13359 16016: contig of 2658 bp in length
* 16017 16116: gap of 100 bp
* 16117 18620: contig of 2504 bp in length
* 18621 18720: gap of 100 bp
* 18721 21036: contig of 2316 bp in length
* 21037 21136: gap of 100 bp
* 21137 23167: contig of 2031 bp in length
* 23168 23267: gap of 100 bp
* 23268 26524: contig of 3257 bp in length
* 26525 26624: gap of 100 bp
* 26625 28953: contig of 2329 bp in length
* 28954 29053: gap of 100 bp
* 29054 32709: contig of 3656 bp in length
* 32710 32809: gap of 100 bp
* 32810 35161: contig of 2352 bp in length
* 35162 35261: gap of 100 bp
* 35262 36494: contig of 1233 bp in length
* 36495 36594: gap of 100 bp
* 36595 40718: contig of 4124 bp in length
* 40719 40818: gap of 100 bp
* 40819 44641: contig of 3823 bp in length
* 44642 44741: gap of 100 bp
* 44742 48769: contig of 4028 bp in length
* 48770 48869: gap of 100 bp
* 48870 53265: contig of 4396 bp in length
* 53266 53365: gap of 100 bp
* 53366 57227: contig of 3862 bp in length
* 57228 57327: gap of 100 bp
* 57328 60723: contig of 3396 bp in length
* 60724 60823: gap of 100 bp
* 60824 64959: contig of 4136 bp in length
* 64960 65059: gap of 100 bp
* 65060 69229: contig of 4170 bp in length
* 69230 69329: gap of 100 bp
* 69330 75479: contig of 6150 bp in length
* 75480 75579: gap of 100 bp
* 75580 84431: contig of 8852 bp in length
* 84432 84531: gap of 100 bp
* 84532 92387: contig of 7856 bp in length
* 92388 92487: gap of 100 bp
* 92488 99945: contig of 7458 bp in length
* 99946 100045: gap of 100 bp
* 100046 112666: contig of 12621 bp in length
* 112667 112766: gap of 100 bp
* 112767 126095: contig of 13329 bp in length
* 126096 126195: gap of 100 bp
* 126196 145450: contig of 19255 bp in length.

```

# FEATURES

```

source
1..145450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-168F19"
/clone.lib="RPC1-11 Human Male BAC"
1..1418
/note="assembly_fragment"
1519..2614
/note="assembly_fragment"
2715..4205
/note="assembly_fragment"

```

```

misc_feature
4306..7004
/note="assembly_fragment"
7105..9239
/note="assembly_fragment"
9340..11600
/note="assembly_fragment"
11701..13258
/note="assembly_fragment"
13359..16016
/note="assembly_fragment"
16117..18620
/note="assembly_fragment"
18721..21036
/note="assembly_fragment"
21137..23167
/note="assembly_fragment"
23268..26524
/note="assembly_fragment"
26625..28953
/note="assembly_fragment"
29054..32709
/note="assembly_fragment"
32810..35161
/note="assembly_fragment"
35262..36494
/note="assembly_fragment
clone_end:SP6
vector_side:right"
36595..40718
/note="assembly_fragment"
40819..44641
/note="assembly_fragment"
44742..48769
/note="assembly_fragment"
48870..53265
/note="assembly_fragment"
53366..57227
/note="assembly_fragment"
57328..60723
/note="assembly_fragment"
60824..64959
/note="assembly_fragment"
65060..69229
/note="assembly_fragment"
69330..75479
/note="assembly_fragment"
75580..84431
/note="assembly_fragment"
84532..92387
/note="assembly_fragment
clone_end:T7
vector_side:left"
92488..99945
/note="assembly_fragment"

```

```

alignment_scores:
    Quality: 466.00      Length: 432
    Ratio: 3.149        Gaps: 5
Percent Similarity: 34.259 Percent Identity: 28.704

```

```

alignment_block:
US-09-512-581-2 x AC068352/rev ..

```

Align seg 1/1 to reverse of: AC068352 from: 1 to: 145450

```

871 SerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVa 887
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
92330 AGTAATCTGATATGCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAAT 92281

887 llLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnT 904
:||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
92280 GAAGCTTGCTCAGGAACCTTGTTACCATGAATATTATCCCCAGAACACT 92231

```



4355 TATATCTGAAGTGTGTTACT.....CTTTGCTGCCATTGA 4389  
 910 .....AlaIleAsnAspGluCysTyrGlnValar 919  
 4390 ATAATATCTGGAATAATGTGCTTTACAGAGTAATGCTATCAAGTAAG 4439  
 919 gGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuP 936  
 4440 ACAAGTGTTCGCCAGAACTTCACAAAGCCCTTCCCGTTTACGGCTTC 4489  
 936 roLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysaspProValLys 952  
 4490 CACTTCAGTATATGCAATCTGTCCTTGTGCAAAAGATCTGTAAAG 4539  
 953 GluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValar 969  
 4540 GAGAGAGAGCTCATGCTAGGCAATGTTGGTGAAATAATAATGTAAG 4589  
 969 gArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerL 986  
 4590 GCGGGAGTATCTGAAGCAGCATCGAGCTGTTAGTGGTAAGCATATAAGAA 4639  
 986 euLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAsp 1002  
 4640 AATGGAAGGATACCTTTTTCAGCCTGCTAGCTTTCAGTTTATAGAATATA 4689  
 1003 ProAspTyrValLysValGlnAspIleGluGlnLysAspValLysG1 1019  
 4690 GCATGATATATG.....ATATTACTG..... 4701  
 1019 ucYsLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnS 1036  
 4702 .....ATATTACTG..... 4710  
 1036 erHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052  
 4710 ..... 4710  
 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCy 1069  
 4711 .....TTATTATTATTG 4724  
 1069 sAspValAlaMetAsnIleIleMetSerLysSerThrTyrSerLeuG 1086  
 4725 TGATTACAAATT..... 4737  
 1086 luserProLysaspProValLeuProAlaArgPhePheThrGlnProAsp 1102  
 4738 .....TTATTACANNNNNNNNNNNNNNNNNNNNNNNNNN 4770  
 1103 LysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPh 1119  
 4771 NNN 4820  
 1119 ePheThrProGlyLysProLysThrThrAsn.....ValL 1131  
 4821 NNN 4870  
 1131 euGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr 1147  
 4871 TCGGCTCTNANAAAAACA.CTTTTCATCAGCAGGCAAGCAATCTCAGACC 4919  
 1148 LysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPr 1164  
 4920 AAATCATCAGATGGAACCTGTGAAGCAATGCAAGCAGCAGCTCAATCC 4969  
 1164 oSerSerProGlyArgIleLysGlyArg 1173  
 4970 AAGCTCTCTCGGAATAAAGGGAGG 4997

seq\_name: gb\_htg14:AC068352

seq\_documentation\_block:

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AC068352 145450 bp DNA HTG 25-JUN-2000  
 Homo sapiens chromosome 4 clone RP11-168F19 map 4, WORKING DRAFT  
 SEQUENCE, 31 unordered pieces.  
 AC068352  
 AC068352.2 GI:8705054  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 145450)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campolano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McSurk,A., McKernan,K., McPheters,R.,  
 Meidirm,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 25, 2000 this sequence version replaced gi:7677723.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L10143  
 Center clone name: 168\_F19

----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 126425 bases at least Q40  
 Consensus quality: 135377 bases at least Q30  
 Consensus quality: 139573 bases at least Q20  
 Insert size: 172000; agarose-fp  
 Insert size: 142450; sum-of-contigs  
 Quality coverage: 3.4 in Q20 bases; agarose-fp  
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
 consists of 31 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1418: contig of 1418 bp in length  
 \* 1419 1518: gap of 100 bp



```

524 .....AlaIlePheSerLy 528
774 TCTGTTGCTTGGGATAAGTTGAACACCTTATAGCCCTTTATAGTTT 2823
528 sValMetVal...IleThrArgAsnLeuProAspPro..... 539
2824 GCTCTCAGCCTGAGTTTCAAGTTTCTCTGCTCCCTATTATGAATCT 2873
540 .....GlyLysAlaGlnAspPheMetLys.....LysPheThr 550
2874 GTGTTTCACTTACTTCGATTACTTGAATTTGTGTAAACAGTCTTTTTC 2923
551 GlnValLeuGluAspGluLysIleArgLysGlnLeuGluValLeuVa 567
2924 TCTGTTGTCAACTATCTCGAAATTCGCTCCT.....CT 2958
567 lSerProThrCysSerCysLysGlnAlaGluGly..... 578
2959 TTCCAGACAAACAACATGAACAACCAAGGACATGTTGACACCTT 3008
579 .....CysValArg.....GluIleThrLysLysLeu 587
3009 TTGAATGCTTTAAATACAGTTTGGGGTTACCACATCACTGTGACCCTC 3058
588 .....GlyAsnProLy 591
3059 TTTCTTTTTTTTTCCTTCTTACGCTGGAAATGTGAGGGTCTGCAGTC 3108
591 sGlnProThrAsnPro...PheLeuCluMetIleLysPheLeuGluVa 607
3109 ACATCCCACTTCTGCGGTTTACTGTCGATTGTCGAGATTACTTAATC 3158
607 rgIleAlaProValHisIleAspThrGluSerIleAlaLeulle... 622
3159 ATTTGTCCTT.....AGTTTCTTATTTCG 3184
622 ..... 622
3185 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3234
622 ..... 622
3235 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3284
623 .....LysGlnValAsnLysSerIleAspGlyThrAlaAsp... 634
3285 ACATAACTGGTGCAGAAACCGCAACTCGTTAGAGAAAACCTGCTGATTT 3334
635 AspGluAspGluGly...ValProThrAspGlnAlaIleArgAlaGlyLe 650
3335 GATTGAGATCATGGGAATTTGCTAGG..... 3361
650 uGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSera 667
3362 .....AAAGTGAAGTTT...CACCCCTTAGATGAACAACA 3398
667 lAcLutThrPheGluSerLeuLeu...AlaCysLeuLysMetAspGluL 683
3399 CTGCTCCATTGACCTCTGCTCATGAAGTGATG..... 3432
683 ysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIle 699
3433 .....TGCGAAGGGAGATT 3447
700 GluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisH 716
3448 GAGAAAAG.....CACTC 3461
716 sLysSerLysLysGly.....ProProArgGlnAlaL 727
3462 TAAACTGAAAAAGAAAAAGAACTACTCTGTTTTCCTCGAAGCAATTTA 3511
727 ys.....TyrAlaIleHis.CysIl 733
3512 AAATGACCAGGCGCACTGCTTTTATAGATGAATATGCTTTTACATGCT 3561
733 ehIsAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluP 750
3562 ACATACAGCATTTCAATGAAAAAAAACCCCTAGATCATCTCTAAATG 3611
750 ro..... 750
3612 ATTGATAAAAGATAAAATATGATACATCAATGGAATACCATCATTA 3661
751 .....LeuHisLysSer...LeuAs 756
3662 CATTATAAATACTAGTAGTACATTAATAATCTTCATAAATATTAAGTAA 3711
756 pProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGly... 771
3712 AACATCAAGATACCTAACTTAATTAATACATAATATTGAAACCTTCTAAA 3761
772 .....HisIle 773
3762 CCACAGCATATACAAAAAGTAGGAACATCTTAGAGTGGCAACATAAA 3811
774 AlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerTrpVala 790
3812 GTGCAATTTCTTCGATTTCATGCTTTCAATAATGCTGAGTTATTGTT 3861
790 lThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
3862 ACACCTTATTATTAAG...CTACCAAAAACTTTGAGTACACAGACCCA 3908
807 LysThrThrLysLeuTrpVal...ProAspGluGluValSer..... 819
3909 CACTCTCTGTAATGTTGGTATTACCTTTTGCAGATATAAACATTTATCA 3958
819 ..... 819
3959 CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4008
819 ..... 819
4009 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4058
820 .....ProGluThrMetValLysIleGlnAlaIl 829
4059 NAATACTATTNGCAAAACCCCCACACTTGCTTANAGAACTCCC..... 4103
829 eLysMetMetValArgTrpLeuLeuGlyMetLys..... 840
4104 .....TATATGAGATGGTGTAAAGCCTTAATTTTGTATCTATTATTA 4146
841 .....AsnAsnHis 843
4147 AGTTAGGCGAGATTATATTTACTAGTTTTTCTCTTTTAAAGTATGT 4196
844 SerLysSerGlyThrSerThrLeuArgLeuLeu..... 854
4197 TTGTTTCTTAACACATCTAACTGGTACTTATTAGGGCCCGCAGATATTT 4246
855 .....ThrThrIleLeuHisSera 861
4247 GCAGCTTCAAAATGAACCTTTAGTCACACATCTCCTCTTATGACGAGTG 4296
861 spGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
4297 AAGCAATTTATCTAGAT..... 4313
878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCy 894
4314 ...AAACTGCTACTGTTTACTCTTTTACTTAAATAGCAAT.....TT 4354
894 styrHisGluIleIleThrLeuGluGlnThrGlnLeuCysAlaLeu... 909

```



```

* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length
* 44741 44840: gap of 100 bp
* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length
* 49510 49609: gap of 100 bp
* 49610 50331: contig of 722 bp in length
* 50332 50431: gap of 100 bp
* 50432 51135: contig of 704 bp in length
* 51136 51235: gap of 100 bp
* 51236 51981: contig of 746 bp in length
* 51982 52081: gap of 100 bp
* 52082 52783: contig of 702 bp in length
* 52784 52883: gap of 100 bp
* 52884 53580: contig of 697 bp in length
* 53581 53680: gap of 100 bp
* 53681 54398: contig of 718 bp in length.
FEATURES
  Location/Qualifiers
    source          1..54398
    /organism="Homo sapiens"

```

```

alignment_scores:
  Quality: 554.50      Length: 1213
  Ratio: 1.160         Gaps: 53
  Percent Similarity: 39.406   Percent Identity: 25.144

```

```

alignment_block:
  US-09-512-581-2 x AC068224

```

```

Align seg 1/1 to: AC068224 from: 1 to: 54398

```

```

283 SerAsnAspAsnGluGluArgLeuGlnValValLysLeuAlaLysMe 299
|||||
1956 AGCAATGATGATGAGGAGCGCTACAAAGTTGTTAAACTACTGGCAAAAT 2005

```

```

299 tPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpG 316
|||||
2006 GTTTGGGGCAAGGATTCAGAAATGGCTTCTCAAAACAAGCCACTTTGGC 2055
|||||
316 lncCystyrLeuGlyArgPheAsnAspIle..... 325
|||||
2056 AGTGCTACTTTGGGCAGGTATATGATTTGGTTTCCCATTTAAAGTAATA 2105
|||||
326 .....HisValProIleArgLeuG1 332
|||||
2106 CATTAATTTCTACTGACCGTTTTTTTAAAGAAACATTATCCA...CATTTAAA 2152
|||||
332 uCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaL 349
|||||
2153 CTGTGTT.....CTCTGGAATCATATAAATGAAGGAT 2184
|||||
349 ysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAla 365
|||||
2185 TTTTATGTTGAAAGTTTGAAGTT..... 2210
|||||
366 lIeArgHisAspValIleValSerIleValThrAlaAlaLysLysAspI1 382
|||||
2210 ..... 2210
|||||
382 eLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal.Arg 394
|||||
2211 .....GTTAATGACGTGGCTATATTATATTACTCTTTTACTGCTTTA 2253
|||||
395 GluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetG1 411
|||||
2254 TCCTCAGTAATTTTAAAGATAATACATTTTAAAGCTATCTCTTATGGT 2303
|||||
411 yLeu.....AlaGlnIleTyrLysTyrAlaLeuGlnSerA 424
|||||
2304 ATTATGACCATGAAGTTAAATATTATTACTTTAAGGCAGTGATTAAATCCA 2353
|||||
424 laAlaGlyLysAspAlaAla..... 430
|||||
2354 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2403
|||||
430 ..... 430
|||||
2404 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2453
|||||
431 .....LysGlnIleAlaTrpIleLys.....AspLysLeuLeuH 442
|||||
2454 NGTCTGNNNGNGSTAACATGGTTAAAGCTGTACCTCGTTANAGTTAAC 2503
|||||
442 isIleTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIle 458
|||||
2504 ACTNCTAC.....TTAGAGGCATTA 2523
|||||
459 PheAlaGlnTyrMetValProHisAsn.....LeuGluTh 470
|||||
2524 TTCATTCTTTTTCCTCTCTCAACCGTCACCAAAATTCGTAAATACT 2573
|||||
470 rThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuA 487
|||||
2574 TTCTCCAAATGCTCTTGAACATTTATCTTCTACTCTCTACCTCTCTATAT 2623
|||||
487 snAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn..... 500
|||||
2624 TTTACCTTGATGTTGTATACATTCCTCCCTGCTGTGTAATGTTCTCTGT 2673
|||||
501 .....Leu.LeuArgHisGlnValLysAspLeuLeuAspLeuI 513
|||||
2674 GCCTCTGACCCACTGCTGCTGTATACATTGCTATCAAAATTTATCTCTC 2723
|||||
513 lelysGlnProLysThrAspAlaSerValLys..... 523
|||||
2724 TAAACACAGATTAGATCATCATCAGTCCCTCTTCTTGAATTTCTGTTGATTC 2773
|||||

```



4621 ACCCTTCCA 4629

seq\_name: gb\_htg14:AC068224

seq\_documentation\_block:

LOCUS AC068224 54398 bp DNA HTG 30-APR-2000  
 DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS

SEQUENCE SAMPLING.

AC068224

VERSION AC068224.1 GI:7671284

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 54398)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 3, clone RP11-660H19

Unpublished

2 (bases 1 to 54398)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
 Boquslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smith, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10161

Center clone name: 660\_H\_19

-----

\* NOTE: This record contains 68 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 \* 686 785: gap of 100 bp  
 \* 786 1513: contig of 728 bp in length  
 \* 1514 1613: gap of 100 bp  
 \* 1614 2354: contig of 741 bp in length  
 \* 2355 2454: gap of 100 bp

2455 3184: contig of 730 bp in length  
 \* 3185 3284: gap of 100 bp  
 \* 3285 3959: contig of 675 bp in length  
 \* 3960 4059: gap of 100 bp  
 \* 4060 4746: contig of 687 bp in length  
 \* 4747 4846: gap of 100 bp  
 \* 4847 5564: contig of 718 bp in length  
 \* 5565 5664: gap of 100 bp  
 \* 5666 6360: contig of 696 bp in length  
 \* 6361 6460: gap of 100 bp  
 \* 6461 7132: contig of 672 bp in length  
 \* 7133 7232: gap of 100 bp  
 \* 7233 7937: contig of 705 bp in length  
 \* 7938 8037: gap of 100 bp  
 \* 8038 8749: contig of 712 bp in length  
 \* 8750 8849: gap of 100 bp  
 \* 8850 9522: contig of 673 bp in length  
 \* 9523 9622: gap of 100 bp  
 \* 9623 10356: contig of 734 bp in length  
 \* 10357 10456: gap of 100 bp  
 \* 10457 11181: contig of 725 bp in length  
 \* 11182 11281: gap of 100 bp  
 \* 11282 11975: contig of 694 bp in length  
 \* 11976 12075: gap of 100 bp  
 \* 12076 12785: contig of 710 bp in length  
 \* 12786 12885: gap of 100 bp  
 \* 12886 13557: contig of 672 bp in length  
 \* 13558 13657: gap of 100 bp  
 \* 13658 14347: contig of 690 bp in length  
 \* 14348 14447: gap of 100 bp  
 \* 14448 15132: contig of 685 bp in length  
 \* 15133 15232: gap of 100 bp  
 \* 15233 15912: contig of 680 bp in length  
 \* 15913 16012: gap of 100 bp  
 \* 16013 16731: contig of 719 bp in length  
 \* 16732 16831: gap of 100 bp  
 \* 16832 17506: contig of 675 bp in length  
 \* 17507 17606: gap of 100 bp  
 \* 17607 18339: contig of 733 bp in length  
 \* 18340 18439: gap of 100 bp  
 \* 18440 19109: contig of 670 bp in length  
 \* 19110 19209: gap of 100 bp  
 \* 19210 19900: contig of 691 bp in length  
 \* 19901 20000: gap of 100 bp  
 \* 20001 20693: contig of 693 bp in length  
 \* 20694 20793: gap of 100 bp  
 \* 20794 21464: contig of 671 bp in length  
 \* 21465 21564: gap of 100 bp  
 \* 21565 22245: contig of 681 bp in length  
 \* 22246 22345: gap of 100 bp  
 \* 22346 23049: contig of 704 bp in length  
 \* 23050 23149: gap of 100 bp  
 \* 23150 23857: contig of 708 bp in length  
 \* 23858 23957: gap of 100 bp  
 \* 23958 24685: contig of 728 bp in length  
 \* 24686 24785: gap of 100 bp  
 \* 24786 25499: contig of 714 bp in length  
 \* 25500 25599: gap of 100 bp  
 \* 25600 26302: contig of 703 bp in length  
 \* 26303 26402: gap of 100 bp  
 \* 26403 27108: contig of 706 bp in length  
 \* 27109 27208: gap of 100 bp  
 \* 27209 27910: contig of 702 bp in length  
 \* 27911 28010: gap of 100 bp  
 \* 28011 28720: contig of 710 bp in length  
 \* 28721 28820: gap of 100 bp  
 \* 28821 29484: contig of 664 bp in length  
 \* 29485 29584: gap of 100 bp  
 \* 29585 30328: contig of 744 bp in length  
 \* 30329 30428: gap of 100 bp  
 \* 30429 31111: contig of 683 bp in length  
 \* 31112 31211: gap of 100 bp  
 \* 31212 31931: contig of 720 bp in length



[illegible]







```

VERSION      L03200.1  GI:168024
SOURCE       DNA-damage sensitive; mitotic block.
SOURCE       Emericella nidulans cDNA to mRNA.
ORGANISM     Aspergillus nidulans
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
             Eurotiales; Trichocomaceae; Emericella.
REFERENCE    1 (bases 1 to 4885)
AUTHORS      Denison,S.H., Kafer,E. and May,G.S.
TITLE        Mutation in the bimd gene of Aspergillus nidulans confers a
             conditional mitotic block and sensitivity to DNA damaging agents
JOURNAL      Genetics 134 (4), 1085-1096 (1993)
MEDLINE      93387663
FEATURES     Location/Qualifiers
             1..4885
             /organism="Aspergillus nidulans"
             /db_xref="taxon:5072"
             133..4653
             /gene="bimd"
             133..4653
             /gene="bimd"
             /codon_start=1
             /protein_id="AA03063.1"
             /db_xref="GI:168025"
             /translation="MPSRTRQGPSVATEDVEPEPSSGGLRQLKFNPLSWRVGRSAIP
             IADLLQRLQTLAQELPNVQEEIEKESLRKVSQELATQAQLLAHKDKGVRAWATCCIVD
             VLRLCAPDAPFTANOLKDIPTCVSIIIPALGDPSPNPNAOHYIIVLSLAEVKSIVLM
             TLDHPDTLIVPLFTSCFDIVAGSAKASTGEPAKNVEYDMTRLLVTVIDESPVLAPD
             VYDVIAQFLRVDPVLDPGPKGKPKETQVDEKQETLLLLKDYPAAYNMAKALCQACP
             EMTSHISQYFNNVIIDASATQNGSPKQARTNLDDSDGEGDIKLSKAHLIRE
             LWRACPDYLVQNIPOLEAELSAESYLRLLATQTIGDLTSGTGAGPPPLPMDPAVY
             POKLDDYARSLPQNPVLLMPFAPKPSQADSHSAYDSFLSRRLDKSASVRASWRTAIG
             RIILTSAGSGSLSDNEOTLIHLSSMLRDLAEVRLAAVEAVGTFLSHIYNKLGVS
             GVSQDLSLLFLILAEVRKDRKSQVREHATKVLARAWAVASGDIERSEHOVTPLLKEAP
             SRILDAYTNDIEIHWISIDRAMFEILLPLSYPIPKLSRSSSSSORLKSQQAPE
             SEADVIRIKVRRLILVGGDEKAKVFAMOKROVSLRTAVTVYLOACEEYNGWYE
             KNKDKIQAQLTKIVDALAKTFDPDARTSADLWKFAKHDRGYQLIRFAMAAYSDYRT
             VTKAKELARLQSSNNTILHETLTLLYRCSIVFNRSHPAIMSIRSDEGLAAP
             AHMKLEIKSEVLEAOVQKDELAQAPKATVSAAGTEELIKACSGFAKKLPS
             KLPKERRKFOALVDVALHSPSPRAAKHVALILMAYDKKMKVAKDLVEKCSKCTVDS
             ERFPLKATLSQNLILAPREADESDAIKISYQILLNLTNRSPFNSGFNSDQVDDDE
             KSKRLWALKIIVNLRUKDGSDDQDFRAHAEPVYDITLNKLINSGELSKKDDTATQ
             KSRLLANLSLLKSSHALCEQLTPODFNSIALVAQDPLPVRSGFINQLKKLV
             OTRLGAWYVYVILIAFEPQVGLKDLSTLWLSRAAFSQOTNGKKGKEQVMEALF
             SBLLSLAYHPDYPPADLDESKLDELDTFARYILFLYSAVANEHNLSTFIHIAORVK
             QARDGTSKDEMSRLHTLSLDAQATIRFADVYSCORFEGGAGVNLQYPPGKVG
             VPSSIFAPNGSHREAEVADTFLPEDAEDRLDRIVRATMRFTKNGSSQAGAKKRKTDS
             TQPSKDAANAARKANSSESSRRKSGSAVGKMPKRRKTKTKGDDSDGEAAGNV
             ASSATRRRSNRGSRRISYADPDSDEDDMEMDLNQARDEDEQDQKIENGSD
             LSELSADSNMLEEPEDDDGPEKEQDPDDKQNGDDDAQAPSPVPVASKAVPGRAMK
             KATLPTRRSARR"
BASE COUNT   1222 a 1320 c 1249 g 1094 t
ORIGIN
alignment_scores:
  Quality: 635.00      Length: 1555
  Ratio: 0.772        Gaps: 58
  Percent Similarity: 52.926  Percent Identity: 20.514
alignment_block:
US-09-512-581-2 x EMBIMD ..
Align seg 1/1 to: EMBIMD from: 1 to: 4885
30 GluMetValArgArgLeuLysMetValVallLysThrPheMetAspMetAs 46
   :::::::::::::::::::::
271 GATTGTGTCAGCGCTGCAACGCTAGCGCAGGAGTGTGCCGACGTCGA 320
   :::::::::::::::::::::
46 pGluAspSerGluGluLysGluLeuTyLysLeuAsnLeuAlaLeuHisL 63
   :::::::::::::::::::::
321 ACAGGAG...GAGATTGAGAAGAGCTCTTAAGGAAGGCTCTCCGAGGAGC 367
   :::::::::::::::::::::
63 euAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeu 79
   :::::::::::::::::::::

```

```

   :::::::::::::::::::::
368 TCGCCAGCGCTCAACTGCTGGCATAAGGATAAGGGCGTTCGGGCCCTGG 417
   :::::::::::::::::::::
80 ValAlaCysCysLeuAlaAspIlePheArgIleTyRAlaProGluAlaPr 96
   :::::::::::::::::::::
418 GCTACTTGTCTCATTTGGTGGATGTGTGGCGCTCTGTGGCTGACGCGCC 467
   :::::::::::::::::::::
96 oTyRThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgG 113
   :::::::::::::::::::::
468 TTTTACGGCG...AATCAACTAAAGATATTTTACTTGCATTGTGTCGT 514
   :::::::::::::::::::::
113 In...LeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyR 128
   :::::::::::::::::::::
515 CGATCATTTCCCGCGCTAGGCGATCCCTCAATCCTTATATATGCTCCAC 564
   :::::::::::::::::::::
129 PheTyRLeuLeuGluAsnIleAlaTrpValLysSerTyRAsnIleCysPh 145
   :::::::::::::::::::::
565 ATTTACGCTTTGAATTCGTCGGGAGGTCAAAAGTATTTGTTCTCATGAC 614
   :::::::::::::::::::::
145 eGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyRArgThrLeuP 162
   :::::::::::::::::::::
615 GGATCTAGTACCCTGGGACACATGATGCTCCCACTGTTTATAGCTGCT 664
   :::::::::::::::::::::
162 heSerValIleAsnAsnGlyHisAsn.....ProAlaHisLysAsnLeu 170
   :::::::::::::::::::::
665 TTGACATTGTGCGAGGCTCAGCTAAAGCCTCGACCGGCGAACCACTGCC 714
   :::::::::::::::::::::
171 GlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGl 187
   :::::::::::::::::::::
715 AAAAACCTCGAATATGATATGACCCGCTGTGTAGTGACAGTATTATGACA 764
   :::::::::::::::::::::
187 uGlyAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuV 204
   :::::::::::::::::::::
765 GTCCGAGCTCTCGCGCTGATGTTGTGGATGATGATGATGATGATGATG 814
   :::::::::::::::::::::
204 al.....ProAlaHisLysAsnLeuAsn 211
   :::::::::::::::::::::
815 TGCCTGTGCTGCTCTCGCGTGTGATGTCGCCCGGAGGAAAAAGAAAAA 864
   :::::::::::::::::::::
212 LysGln..... 213
   :::::::::::::::::::::
865 CCCGAGACTCAGTGGAGGAGAAACAAGAGACGCTGCTACTAAAGGACTA 914
   :::::::::::::::::::::
214 .....AlaTyRAspLeuAlaLysAlaLeuLysArgThrAlaGlnA 228
   :::::::::::::::::::::
915 CCCGCTCGCTACACATGCGCAAGCAATCTGCGAGGCTTGCCTCGGAGA 964
   :::::::::::::::::::::
228 laIleGluProTyRThrPhePheAsnGlnValLeuMet..... 242
   :::::::::::::::::::::
965 GGATGACTAGTCATATCAGCAGCTACTTTAAACAAGCTCATAATTGACGC 1014
   :::::::::::::::::::::
243 .....LeuGlyLysThrSe 247
   :::::::::::::::::::::
1015 TCAGCCACGGGAACCAAAATGGGCCCTCAAGCAAGCTCGCAGAACTAA 1064
   :::::::::::::::::::::
247 rIleSerAspLeuSerGluHisValPheAsp..... 257
   :::::::::::::::::::::
1065 CCTCGATGACTCGGACGAGGAGGAGGAGACATCAAGGAATTGAGTAAAG 1114
   :::::::::::::::::::::
258 .....LeuIleLeuLeuTyRAsnIleAspSerHisLeuLeuLeu 271
   :::::::::::::::::::::
1115 CGCATCGATTGATCCGTGAGCTGTGGAGAGCATGCCCGGACGTTTTCGAG 1164
   :::::::::::::::::::::
272 SerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspGluGl 288
   :::::::::::::::::::::
1165 AATGTTATCCCCCAGCTAGAGAGCAACTGTCCGCGCAATCGGTGTCATT 1214
   :::::::::::::::::::::
288 uArgLeuGlnValVallLysLeuLeuAlaLysMetPheGlyAlaLysAsps 305
   :::::::::::::::::::::
1215 CGCGTTG.....TTAGCAACCCAGACCATCGCGGACCTGACAT 1252
   :::::::::::::::::::::
305 er.GluLeuAlaSerGlnAsnLysPro.....LeuTrp.GlnCyst 318
   :::::::::::::::::::::

```



```

891 GlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCy 907
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3722 ACACAGAAGCATTTTCACGATATGCTCAGCCATCATAGATTTCAACCTTT 3771
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
907 sAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaG 924
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
3772 GGCCTTACTACCAAGATGATGTCAGAGAGGTGCTCAGGTTTGTGTC 3821
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
924 InLysLeuHisTysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
   ::::: ::::: ::::: ::::: ::::: ::::: :::::
3822 GGAAGCTCCAAAAGTATCTGCGCGACGGAAGCTGCGGTCTACTAC 3871
   ::::: ::::: ::::: ::::: ::::: ::::: :::::
941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHi 957
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
3872 ACCATCATCTTCCTCACTGCTTCGACACCGTCGAGTCAAG..... 3916
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
957 sAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuL 974
   ||| ||| ||| ||| ||| ||| ||| ||| |||
3917 .....AACAGGTGGAGACATGGATCC 3938
   ||| ||| ||| ||| ||| ||| ||| ||| |||
974 ySlnHisAlaValSerGluLysLeuSerLeuLeuProGluTyr 990
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
3939 GTTCCGAGCGGCCACTTCCAGAACCTCAAGCAGCTGTCTCGAGCG 3988
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
991 ValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVally 1007
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
3989 ATCATGGCTCGCTGATCTCGTTCCTGGGCCACACCCCGACTAC..... 4033
   ||| ||| ||| ||| ||| ||| ||| ||| |||
1007 sValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheV 1024
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
4034 .AGCAATAGGTGGAAATACCTCGTCGACCAACCGCCCG.....TACA 4073
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
1024 alLeuGluLeuLeuMetAlaLysAsnGluAsnSerHisAlaPheIle 1040
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
4074 TCCTCAACTACATAGTGGTGGTAGGCCACCGAGTCCAAATTTGGGCTTGATA 4123
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAs 1057
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4124 TACAAGTACCGGAGCGGTGTGAAGCAGACCAAGATGGACTGAACCCGAA 4173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1057 p...AspAla.....L 1060
   : |||||
4174 CAGCGAGCCGCCCGCTTGTAGTGACTGGCGCAATCATATCAGGA 4223
   || ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 yMetAsnGluLys..... 1064
   || ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
4224 AGTGGCAGGAGAAAAGTTTGGGCTTTCGCTCGTTCGCCGCAAGGTC 4273
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1065 .....LeuTyrThrValCys.....AspValAlaMe 1073
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
4274 GGCCTACCTCTCGGCTGTACAGGCCCTTCACTCCACAGTCAAGCGCA 4323
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1073 tAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysA 1090
   : ||| ::::: ::::: ::::: ::::: ::::: :::::
4324 GCGCATCGCGGAAGATCGTATCTCCCGAGGCGCTTGGATGAGAAGCTGS 4373
   : ||| ::::: ::::: ::::: ::::: ::::: :::::
1090 spProValLeuProAla.....ArgPhePheThr 1099
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4374 ACAGAGTCTCAGGGGTATGGATCGGAAGAAGTACGAGAGGAAACT 4423
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1100 GlnProAspLysAsnPheSerAsnThr.LysAsn...TyrLeuProProG 1115
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4424 CGACCGGACCTCGTACTAGTACAGCCCTTCACTACGTCCTATTCCTTTGT 4473
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1115 LuMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeu 1131
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4474 CACAGAAACGAAGTCCACGTCGCGGCCCGCCGCCACCAACACCGCGAC 4523
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSer..... 1145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4524 GGCCATAAGTCCAAAGAAAGTCAAAATCATCGGTGACCGCGAGCGGACGG 4573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1146 .GlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerS 1162

```

seq\_name: gb\_pl3:EMEBIMD

seq\_documentation\_block:

LOCUS EMEBIMD 4885 bp mRNA

DEFINITION Emericella nidulans blocked in mitosis D (bimD)

ACCESSION L03200

PLN

26-OCT-1993

mRNA, complete cds.



1962 CGGCTGAGGAGTCGGACTCAGCGCGGAGGACGAGGCTACTCTGATCCGA 2011  
 350 AspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaI 366  
 2012 GGGTTGGGAGAGAGCTCTCT.....GACAGTATGATAAAGGT 2049  
 366 eArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleL 383  
 2050 CAGACTTCAGCTGTTAAAGCTGTGAGAGACCTTCAAGTTCAGGATATCA 2099  
 383 euLeu.....ValAsnAspHisLeuLeu 390  
 2100 TTGCCAAGCTAGGACCTAATGGAGGAGTACGCAAGGATGGTCTGCTCTC 2149  
 391 AsnPheValArgGluArgTyrLeuAspLysArgTrpArgValArgLysG1 407  
 2150 AACACGTTGGCCGATCGCTGTCCGCGAGGAAACCCCGGTTAGAGTTGC 2199  
 407 uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerA 424  
 2200 CGCCATGTCAGTGTGGCTAAACTCTGGGCTGTCCGGACTGGAGAGATGC 2249  
 424 laAlaGlyLysAspAla....AlaLysGlnIleAlaTrpIleLysAspLys 439  
 2250 TTGCTGAAATGAAGCGTAACAGCAGCGTTGTCTGGCATCCCGTCACGC 2299  
 440 LeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValG1 456  
 2300 ATCTACAAACGCCCTTTACGCCAATGACCTGGAGTGAACCTCTTTGATGGA 2349  
 456 uArgIlePheAlaGlnTyrMetValPro..... 465  
 2350 TCGGCTCATCTATCAGTTCCTGGTGCTCTCTGGGTTTCCGCCACCCAGA 2399  
 465 ..... 465  
 2400 AGCGACACGAGGAATTCCAAAGCCCAACGCAACTCCCAATCGCAATCGCG 2449  
 466 .....HisAsnLeuGluThrThrGluArgMetLysCy 476  
 2450 AATCGCGCATCCATCGACCATGATGCTATCAGGCGCTGAGAG..... 2491  
 476 sLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuA 493  
 2492 .ATCCTCTCTGCGCGCTCCTCAGCAGCGCGCGAAGAGCATCT 2540  
 493 snGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509  
 2541 TCGGCATCGAGTCTAGACGGCTCAGTTCCGCAAGATTTTGGAAACCTAC 2590  
 510 LeuAspLeu.....IleLysGlnProLysThrAs 519  
 2591 CTCGACACGATCGCATCGGTACACGGTGGTGTATGGAAGCAACGCCGA 2640  
 519 pAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnL 536  
 2641 TAAATCACCAGCAACTGTACAAACAGCGCGGATATCATCGCACAGTTCC 2690  
 536 euProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnVal 552  
 2691 TTCCAGAACACGTTAAGAGCAAGACGACCTGATCAAGTTTGGCAAGATT 2740  
 553 LeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuLeuValSerPr 569  
 2741 .....CACGATCGCCCAATTAACCTGATCAAAATATGTCATTGGCCA 2784  
 569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysL 586  
 2785 GGAGAACTGACTTAAAGCGGTTTATAGGCTTTGAAGAGCTTATCAAGC 2834  
 586 ysLeuGlyAsnProLysGlnProThrAsnProPheLeuLeuMetIleLys 602  
 2835 GCTGTATGGCCAGCAAGACCCCTCA.....GTTATCGACACTCTGCTT 2878

603 PheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSe 619  
 2879 CTTTGGCTCTACCGTCCGGTTGCTTCTTCAACCCGACCATCTGTGTC 2928  
 619 rAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspG 636  
 2929 TAGCATCATGAATACTCCAATCAGACAAGGAGCGGCTGGGATCG.... 2974  
 636 luAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeu 652  
 2975 .....ATTGGCGACGAGATC 2989  
 653 LeuLysValLeuSerPheThrHisProIleSerPhe.....HisSerAl 667  
 2990 TTGAACGAGATCTCTCAACCAACCCGACCTATTCAAAACTCATATCGG 3039  
 667 aGluThrPheGluSerLeuLeu.....AlaCysLeuLysMetA 680  
 3040 ACAGCTTTGCAAGGATCTTGTGACCAGGCCCTTACTGCCACAAGCCGA 3089  
 680 spAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly 696  
 3090 ACGATCCCATAGTAGCGAA...ACGCTCAAGGCTTGTCTCCACGTATGCC 3136  
 697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProVa 713  
 3137 CGGAAGTTCCTTAGGATGTCGCCATCGGATCGGAAGTTCGTGACACCAT 3186  
 713 lLeuHisLysLysSerLysLysGlyProProArgGlnAlaLysTyrAlaI 730  
 3187 GATCACTACGCGCTTTATGCCAACCCGTCAGGCTTCGAAACATCTGTG 3236  
 730 leHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAla... 745  
 3237 TCAACATTGTTCTTTGAAGCAGGATGACAAGAGCATGGTCACCTGCCACT 3286  
 746 GlnIlePheGluProLeuHisLys.....SerLeuAspProSerAsnLe 760  
 3287 GACATTCCTCAGCGGATATTGAAGGGTTGGTCGTATGGTCGAGTAAC... 3334  
 760 uGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuA 777  
 3335 .....TTCTTAACAAGCTTACCGCTGTGAGCCAGTTGGAGCTTCTTG 3377  
 777 laProAspGlnPheAlaProTrpLysSerTrpValAlaThrPheIle 793  
 3378 CACCC...AAGGTTACTGAGGATGCCAGTGTAGATGATGATCCTCAATCTCACC 3424  
 794 ValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLy 810  
 3425 TTTAAGGAGATTTCTTCCAGGTCGCCAGCCGACCCCAAGACACGCCGCC 3474  
 810 sLeuTrpValProAspGluGluValSerProGluThrMetValLysIleG 827  
 3475 CGATTGGTCAATGGCCCTGATATGAGGAGGAATCCAGGCCAAGTCC 3524  
 827 lnAlaIleLysMetMet.....ValArgTrpLeuLeuGlyMetLys 840  
 3525 TCTCAGTGAATAATACTAGTCAATAGGGAAGGAGCATCGAGGGTATCGAG 3574  
 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrI 857  
 3575 GAGCAAGGAGGAAGGCC...AGTTCTGTTTGAAGGGTGTCTCAGGAAGAT 3621  
 857 eLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProA 874  
 3622 CATCAAGGAGAGAGGCGAGATGTCGAGGAGAGGAAGAAACGCCAAGCATC 3671  
 874 spMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890  
 3672 ACAAGACGAGGCTGAGGCTTCTGCGCGCACAGTGTGATGCTGAAGCTCTGC 3721



SRDDSPGAAPVESERRQSSRRRAANKAKGPTYTERDSDDDDEMLEGVAWEYLDDEDD  
 EDKEDDESGEAPRRTAAKASATRRSRREDKSEEEVNEEPKQEEBEPADPE  
 QEEAPTAARAKRTSATAGKKAASLVRNAAASRRSGRKAASAAAAAASAKKDDSD  
 LSELSVMWDEPEAEAEAEAEVEAEVEDEKEOEAPPARANGRKAKAAAPAK  
 AKAAPGKAKAADASAPARRSGRSTRGKQOEPEDEDEDAMDEE"  
 894...955

intron /gene="SP076"  
 /number=1

exon /gene="SP076"  
 956...4405

intron /number=2  
 4406...4478

exon /gene="SP076"  
 /number=2  
 4479...5649

polyA\_site /gene="SP076"  
 /number=3  
 5649

polyA\_site /evidence-experimental  
 5887

polyA\_site /evidence-experimental  
 1642 a 1734 c 1718 g 1317 t

alignment\_scores:  
 Quality: 697.50 Length: 1652  
 Ratio: 0.819 Gaps: 60  
 Percent Similarity: 51.574 Percent Identity: 21.247

alignment\_block:

US-09-512-581-2 x SMAJ9934 ..

Align seg 1/1 to: SMAJ9934 from: 1 to: 6411

13 IleThrTyProGlyValLysGluIleSerAspLysIleSerLysGlu 29  
 639 TTGACATGGCCGCCGCCG.....AGCCCATTTTCGGTTGA 673  
 29 uGluMetValArgLeuLysMetValValLysThrPheMetAspMeta 46  
 674 TACACTGCTTAAGAGACTCGACGCCCTCTCCAGGAGCTGTCGGATATGG 723  
 46 spClnAspSerGluGluGluLysGluLeuTyLeuAsnLeuAlaLeuHis 62  
 724 ACCAGGAGACTGTGGACACCGCATTCGCTG...GTCAAGGTCGCTAAGGAC 770  
 63 LeuAlaSerAspPheLeuLysHisProGlyLysAspValArgLeuLe 79  
 771 GTGCGCTCTCACAGCTTATCCAAACAAGGACAAAGGGCGTACGACATA 820  
 79 uValAlaCysCysLeuAlaAspIlePheArgIleTyAlaProGluAla 96  
 821 CACAGCATGCTGATGTGATATCTTAAGACTATGCGCCCGCGATGCGC 870  
 96 roTyThrSerProAspLysLeuLys..... 104  
 871 CCTTTACC...CCATCACAACCTCAAGGTTTGAGAGCAAAAAGATGCCAT 917  
 105 .....AspIlePheMetPheIleThrArg..Gln..... 113  
 918 TTTATGGACCATCTGCTGATCGGTGGCCCGACACAGGACATCTTTAAT 967  
 114 .....LeuLysGlyLeuGluAspThrLysSerProGlu 124  
 968 CTCAGCATAACTCCATCATCTCCGCCCTTTTCGATCCCTCGATCCCTA 1017  
 124 nPheAsnArgTyPheTyLeuLeuGluAsnLeuAlaIlePheValLysSer 141  
 1018 CAACAACACGACACAAGTATGTGTACGGTCTTTTTCGAGATTAAGAGTA 1067  
 141 yrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeu 157

1068 TCGTTCTGCTACTCGATGTCGAAGCGACGCGCCTTGTACTCAAGCTT 1117  
 158 TyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLys..... 172  
 1118 TTTACCACAATTTTCGACGGCTCTCAGGTCTCAAAATCTCCAAGGAGA 1167  
 173 .....ValHisMetHisMetValAspLeuMetSerSerI 184  
 1168 GCAGGTTGGAAGGACGTCGAGTTTAGTATGCAAGAGATGTTGGGAGCGC 1217  
 184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200  
 1218 TCATTGACGACTCCGTGACCCCTTCCAGGAAGGTGCTCGATGTCATCATG 1267  
 201 ValAsn...LeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyAs 216  
 1268 GCCAGTTCTCGAGACTGCTGCTCGGGACTAGGCAAGGACGCCAGGA 1317  
 216 p.....LeuAlaLysAlaLeuLeuLysArgThrA 226  
 1318 TCATGTGCCAATCGACGACGACGCAACACTGCTCCTGAAGGAAGAGC 1367  
 226 laGlnAlaIleGlu..... 230  
 1368 CTGAGGCTTACCAGATGTCAGGAATCTATGCCAGACTTATGACGACAAG 1417  
 231 .....ProTyThrPhePheAsnGlnValLeuMet..... 242  
 1418 ATGGCCCGTTTCGCCAGTCAATATTTTAGTGCCTGATGTCGACCGGAC 1467  
 243 .....LeuGlyLys..... 245  
 1468 AGTGTGTGGGAAGCGAAGTGTAGCAGACGACGAAGATGACGAAAACG 1517  
 246 .....ThrSerIleSerAspLeuSerGlu... 253  
 1518 ATGAGGCTGATGGCCCAACAGGCCCTTCGAGTCCGACTTGAAGGAATTG 1567  
 254 .....HisValPheAspLeuIleLeuGluLeuTyAsnIleAspse 267  
 1568 AGGAAGCGCATGTG.....CTTATTCGGAGACTCTGGAAGCTCCCC 1611  
 267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSera 284  
 1612 CATGATCCTTCAAAAGCTTATCCCCAGGTTGATCGAGAGCTCTCGGCAG 1661  
 284 snAspAsnGluGluArgLeuGlnValValLysLeuAlaLysMet... 299  
 1662 ACAATGTCACCTTGGCAGACAGATGGCAACGAGACCCCTCGGACATGATT 1711  
 300 .....PheGlyAlaLys..... 303  
 1712 TCAGGCAATTGGCGCGCAGGTCCGCCCTCCACTGCCCATCTTGTATCCAGC 1761  
 303 ..... 303  
 1762 TGCTWACCGCCATTGAGCCTGGAAGAGGAAGATAGAGCGGAACCCCCAG 1811  
 304 .....AspSerGluLeuAlaSerGln 310  
 1812 TGACCAACATCTCACCGCGCTTGTCTCCATCTCTTCGCGACAGA 1861  
 311 AsnLysProLeuTrpGlnCysTyLeuGlyArgPheAsnAspIleHisVa 327  
 1862 CATAGCAGCAGCTTCCACAACCTTCTGAGTAGAAGAATGATGAAGGCTCC 1911  
 327 lProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMet... 342  
 1912 CTCTATTGCGCGATGGACCGCAGTAGGCCACATTTCTCTACTT 1961  
 343 .....AsnHisProAspLeuAlaLys 349  
 ..... ||| :::



203 AACCAGAGATGCCCCAGTCTCCAGATGAATCAAGACAAATCAAAACTGT 252  
1066 yrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThr 1082  
253 ATACGATGTGATGGCTCTCTGTGTATATAAATAGTAAAGTCTTG 302  
1083 TyrSerLeuGluSerProLysAspProValLeuProAlaArgPheTh 1099  
303 TGCATCCAGATCAACCAAGGCCAGTCTCCCAATGAAATTTTAC 352  
1099 rGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluM 1116  
353 ACAACCTGAAAGGACTCTGTAAAGGATAAGAGTTATATTCAGACAGA 402  
1116 etLysSerPheThrProGlyLysProLysThrAsnValLeuGly 1132  
403 CAAGAGTACTTCTGTTAACAGGAAGCAAGGCTGCTGGAGTACTAGT 452  
1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSe 1149  
453 GCAGTAATAAGCTTTATCAGCACCGGGAAGGAACCTATGTAGAAG 502  
1149 rSerArgMetGluThrValSer.....AsnAlaSerSerSerAsnP 1164  
503 CACTGGCACTGAGACTGGAAGCAATATTAATGTAATTCAGAGCTGAACC 552  
1164 roSerSerProGlyArgIleLysGlyArg...LeuAspSerSerGluMet 1179  
553 CTCAACCGGAAATTCGATCAAGGACAGAGTTTCAGAGCGCAGCAAACT 602  
1180 AspHisSerGluAsnGluAsp.....TyrThrMetSerSerProLeuPr 1194  
603 GGAGTTAGTGAATGAAGAACCCCTGTGAGGATTTATTCAGTCACACC 652  
1194 oGlyLysSerSerAspLysArgAspSerAspLeuValArgSerGluL 1211  
653 TGTAAGAATAATTGAC..... 668  
1211 euGluLysProArgGlyArgLysLysThrProValThrGluGlnGlu 1227  
669 ..... 686  
1228 LysLeuGlyMetAspLeuThrLysLeuValGlnGluLysProLy 1244  
687 ...ATTAATTCATCAGCTACC.....CAGGCAACAT 718  
1244 sGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspG 1261  
719 CAGCAGTACCCGAGGAAAGAAAGAACAGTAAACAGCAGCTGTCGACAGA 768  
1261 lu...GlnGlnTrpProGluLysArgLeuLysGluAspIleLeuGlu 1276  
769 ATATCCCAACAAAAACAGATGAGAAA..... 794  
1277 AsnGluAspGluGlnAsnSerPro...ProLysLysGlyLysArgGlyAr 1292  
795 ...GTAGATGATCGGACCTCCCGCCCTCCCAACCCAGGAGGAGC 841  
1292 gProProLysProLeuGly...GlyGlyThrProLysGluGluProThrm 1308  
842 TCGACCAAGTCTGAATCTCAGGCAATGCTACCAAAAAATGATGATCTAA 891  
1308 etLysThrSerLysGlySerLysLys...SerGlyProProAlap 1324  
892 ATAAACCTATTAAACAGGGAAGGAGAGCTGCGAGTGGGTCAGGAGC 941  
1324 ro.....GluGluGluGluGlu 1331  
942 CTGGGGTGGAGCAGGTAAATGCCAACACCACTGCAAGATT 991  
1332 ArgGlnSerGlyAsnThrGluGlnLys.....SerLysSerLy 1344  
992 AGCCAAAAAGGCAGCACCAGCAGAGAAACAAATTTGACTTACAAAGGTAAA 1041

1344 sGlnHis.....ArgValSerArgAlaGlnGlnArgAlaGluSerp 1359  
1042 AATGATTCGAAGGAGAGAAATGAAGGCCAAACAGAAG...CAGGCTC 1088  
1359 roGluSerSer 1362  
1089 CAGCTTCTGCA 1099  
seq\_name: gb\_pl4:SMJ9934  
seq\_documentation\_block:  
LOCUS SMJ9934 6411 bp DNA PLN 30-JUL-1999  
DEFINITION Sordaria macrospora SP076 gene.  
ACCESSION AJ009934  
VERSION AJ009934.1 GI:5679711  
KEYWORDS SP076 gene.  
SOURCE Sordaria macrospora.  
ORGANISM Sordaria macrospora.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariales; Sordariaceae; Sordaria.  
REFERENCE 1 (bases 1 to 6411)  
AUTHORS van Heemst,D., Poggeler,S. and Zickler,D.  
TITLE The SP076/PDS5/BIND gene involved in mitotic sister chromatid cohesion is also needed for synaptonemal complex formation and recombination in Sordaria macrospora  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6411)  
AUTHORS van Heemst,D.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUL-1998) van Heemst D., Agricultural University, Department of Genetics, Dreijenlaan 2, 6703HA Wageningen, THE NETHERLANDS  
FEATURES  
source  
1..6411  
/organism="Sordaria macrospora"  
/db\_xref="taxon:5147"  
mRNA join(138..893,956..4405,4479..5649)  
/gene="SP076"  
gene 138..5649  
/gene="SP076"  
exon 138..893  
/gene="SP076"  
/number=1  
CDS join(543..893,956..4405,4479..5648)  
/gene="SP076"  
/codon\_start=1  
/product="Spo76 protein"  
/protein\_id="CA51808.1"  
/db\_xref="GI:5679712"  
/translation="MAPRSAPPEEELEEQEQGLVALSEPLTWRPKPTISVD  
TLRLKRLDLSKLSMDQETVDTSLVKAKDVASHQLIHKDKGVRATACCI  
RLCAPDPTPSQLKIDFNLSINIIIPALFDPSPYNNQHKYVLSFAIKI  
VEGSEALILKFTTIFDVGSLKSSGQGVKDFEFSMOEMGALIDDSVTLP  
VIMAFVLAADGPKGERODHPIDDOATLLKKEPEAYQVRLNCOYDQDKM  
SOEVSIVDAATGACPKNGSROEDDDEGPGTGPSPESDLKELKWLREI  
AEMILQNIPOVDAELSDNVHLRQMATETLGDMSGIGAGPPPLPILDP  
LEEDRAEPPVTNLTPLICSISFSQTHSTTFHNSLRNDRKAPSIRAWT  
SATSAGIGLSREDEATLIRGLGKLSDSDEKYLRSKAVETFKFODITAK  
GKDSVINTLADRCRDRKPAVRAAMSLLAKLWAGTVKNSNANGSOANA  
YNAFVNLDELAVLMRDVIYFLPLGPPPAKATRNNSANGSOANAIDHDA  
IRAERTILLARSIDRPAKFAFFAMOSRRPQFAKILETYLDQCDRYNG  
SNLNTADYIAQFLPEHVKSTDLKFAKHBRNNYLLIKYVIGENDKTYK  
LIRCMASKDPSVIDTLLPLLYRSGCLLENRSHLSIMEYSKDKDGLSIA  
ISORNPDLFTGHIGKCLDLPQATPNPDIVAETLKACSTYARKPKDVM  
FYOTMINTALYQGPVKASHAVNVLCKQDDKSMVTATDILQILKNGSY  
LTAVSOLELLAPKVTEDASDEILNLTFFKILLQVTRDAKSDPDWNGA  
CLSLKILNVRVSIIEGIEAKESKSSVKVLKIIKEKGEIVEEKEPKH  
AAQLMLKCTQKHDDMLTPSDFNLLALTQDMVEVRGFRKKLQYLADK  
YTIIFLTAPEPSVEFKNRVETWIRSRARFONLQPVLEAIWRLISLHA  
VEYLDHARYILNYIVLGTESNLGLIYKIAERVQKOTDGLNPSDAHVL  
IRKQEKKWFAAFAPGKVLPLGLYTLAQSHSEQAIAEKSYLPGLKDEL  
MDRKKRKRKSTSAATGTNNRDGHKSKVKSSGDRGDRKSSRKIATAKSK



\* 53814 53913: gap of 100 bp  
 \* 53914 54675: contig of 762 bp in length  
 \* 54676 54775: gap of 100 bp  
 \* 54776 55537: contig of 762 bp in length  
 \* 55538 55637: gap of 100 bp  
 \* 55638 56413: contig of 776 bp in length  
 \* 56414 56513: gap of 100 bp  
 \* 56514 57278: contig of 765 bp in length  
 \* 57279 57378: gap of 100 bp  
 \* 57379 58149: contig of 771 bp in length  
 \* 58150 58249: gap of 100 bp  
 \* 58250 59023: contig of 774 bp in length  
 \* 59024 59123: gap of 100 bp  
 \* 59124 59908: contig of 783 bp in length  
 \* 59907 60008: gap of 100 bp  
 \* 60007 60792: contig of 786 bp in length  
 \* 60793 60892: gap of 100 bp  
 \* 60893 61663: contig of 771 bp in length  
 \* 61664 61763: gap of 100 bp  
 \* 61764 62539: contig of 776 bp in length  
 \* 62540 62639: gap of 100 bp  
 \* 62640 63422: contig of 783 bp in length

## alignment\_scores:

Quality: 768.00 Length: 147  
 Ratio: 5.224 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-512-581-2 x AC016449/rev ..

Align seg 1/1 to reverse of: AC016449 from: 1 to: 72157

1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluG1 1225  
 39685 TCTGAATTGGAGAGCCCTAGAGCGAGGAAAAACCCCGTCACAGAAC 39636  
 1225 nGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluInL 1242  
 39635 GGAGGAGAAATTAGGTATGTGATGACTTGACTAAGTTGGTTACAGGAACAGA 39586  
 1242 yProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258  
 39585 AACCTTAAGGCACTCAGCGAAGTCCGAAAGAGGCCATACGGCTTCAGAA 39536  
 1259 SerAspGluGlnGlnTrpProGluLysArgLysGluAspIleLe 1275  
 39535 TCTGATGAACACAGCTGGCCCTCAGGAAAGAGGCTCAAGAGAGATATATT 39486  
 1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA 1292  
 39485 AGAAATGAAGATGAACAGAAATAGTCCGCCAAAAAGGGTAAAGAGGCC 39436  
 1292 rGProLysProLeuGlyGlyThrProLysGluGluProThrMet 1308  
 39435 GACCACCAAAACCTCTTGCTGGAGGTACACCAAGAGAGAGCCACCAATG 39386  
 1309 LysThrSerLysGlySerLysLysLysSerGlyProProAlaProG1 1325  
 39385 AAAACTTCTAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGA 39336  
 1325 uGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerL 1342  
 39335 GGAGGAGGAAGAAGAAAGCAAGTGGAAATACGGAACAGAAAGTCCA 39286  
 1342 ySerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355  
 39285 AAAGCAACAGCACCAGTGTCTAAGGAGAGACACAGCAGAGG 39245

seq\_name: gb\_pr5:AK021757

seq\_documentation\_block:

LOCUS AK021757 2212 bp mRNA PRI 29-SEP-2000

## DEFINITION

Homo sapiens cDNA FLJ11695 fis, clone HEMBA1005019, highly similar to Homo sapiens mRNA for KIAA0648 protein.

## ACCESSION

AK021757

## VERSION

AK021757.1 GI:10433003

## KEYWORDS

oligo capping; fis (full insert sequence).

## SOURCE

Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone\_lib:HEMBAL clone:HEMBAL005019.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (sites)

## AUTHORS

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahari, K., Masubo, Y. and Oshima, A.  
 NEDO human cDNA sequencing project

## TITLE

Unpublished (2000)

## JOURNAL

2 (bases 1 to 2212)

## REFERENCE

Isogai, T. and Otsuki, T.

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert construction; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

## FEATURES

## source

1. .2212  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEMBAL005019"  
 /clone\_lib="HEMBAL"  
 /dev\_stage="embryo, 10 weeks"  
 /tissue\_type="whole embryo, mainly head"  
 /note="cloning vector: pME18SFL3"

## BASE COUNT

751 a 385 c 430 g 646 t

## ORIGIN

## alignment\_scores:

Quality: 706.50 Length: 404  
 Ratio: 2.656 Gaps: 15  
 Percent Similarity: 65.842 Percent Identity: 42.822

## alignment\_block:

US-09-512-581-2 x AK021757 ..

Align seg 1/1 to: AK021757 from: 1 to: 2212

983 LeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLe 999  
 3 TTATTATCATCTGTCCTGAATATGTTAGTTCCATCATGATTCACCTGCT 52  
 999 uAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysA 1016  
 53 ACCCATCATCCAGATTTTACAAAGATCACAAGATGTTGATCAGCTTCGG 102  
 1016 spValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032  
 103 ATATCAAGAGTGCCCTATGTTTCATGCTTGAAGTTTAAATGACAAAGA 152  
 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysG1 1049  
 153 GAAACAAATAGCCATGCCCTTTATGAAGAGATGGCAGAGACATCAAGTT 202  
 1049 nThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeu 1066  
 |||||



All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L1189

Center clone name: 45\_L14

-----

\* NOTE: This record contains 83 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

```

1      757: contig of 757 bp in length
*      758 857: gap of 100 bp
*      858      1635: contig of 778 bp in length
*      1636 1735: gap of 100 bp
*      1736      2490: contig of 755 bp in length
*      2491 2590: gap of 100 bp
*      2591      3343: contig of 753 bp in length
*      3344 3443: gap of 100 bp
*      3444      4224: contig of 781 bp in length
*      4225 4324: gap of 100 bp
*      4325 5088: contig of 764 bp in length
*      5089 5188: gap of 100 bp
*      5189 5968: contig of 780 bp in length
*      5969 6068: gap of 100 bp
*      6069 6844: contig of 776 bp in length
*      6845 6944: gap of 100 bp
*      6945 7701: contig of 757 bp in length
*      7702 7801: gap of 100 bp
*      7802      8573: contig of 772 bp in length
*      8574 8673: gap of 100 bp
*      8674      9430: contig of 757 bp in length
*      9431 9530: gap of 100 bp
*      9531 10295: contig of 765 bp in length
*      10296 10395: gap of 100 bp
*      10396      11139: contig of 744 bp in length
*      11140 11239: gap of 100 bp
*      11240 11988: contig of 749 bp in length
*      11989 12088: gap of 100 bp
*      12089 12874: contig of 786 bp in length
*      12875 12974: gap of 100 bp
*      12975      13750: contig of 776 bp in length
*      13751 13850: gap of 100 bp
*      13851 14638: contig of 788 bp in length
*      14639 14738: gap of 100 bp
*      14739 15520: contig of 782 bp in length
*      15521 15620: gap of 100 bp
*      15621 16381: contig of 761 bp in length
*      16382 16481: gap of 100 bp
*      16482 17248: contig of 767 bp in length
*      17249 17348: gap of 100 bp
*      17349 18123: contig of 775 bp in length
*      18124 18223: gap of 100 bp
*      18224      18977: contig of 754 bp in length
*      18978 19077: gap of 100 bp
*      19078 19860: contig of 783 bp in length
*      19861 19960: gap of 100 bp
*      19961 20732: contig of 772 bp in length
*      20733 20832: gap of 100 bp
*      20833      21590: contig of 758 bp in length
*      21591 21690: gap of 100 bp

```

```

*      21691      22446: contig of 756 bp in length
*      22447 22546: gap of 100 bp
*      22547 23333: contig of 787 bp in length
*      23334 23433: gap of 100 bp
*      23434      24212: contig of 779 bp in length
*      24213 24312: gap of 100 bp
*      24313      25093: contig of 781 bp in length
*      25094 25193: gap of 100 bp
*      25194 25958: contig of 765 bp in length
*      25959 26058: gap of 100 bp
*      26059 26839: contig of 781 bp in length
*      26840 26939: gap of 100 bp
*      26940      27705: contig of 766 bp in length
*      27706 27805: gap of 100 bp
*      27806 28568: contig of 763 bp in length
*      28569 28668: gap of 100 bp
*      28669      29419: contig of 751 bp in length
*      29420 29519: gap of 100 bp
*      29520 30311: contig of 792 bp in length
*      30312 30411: gap of 100 bp
*      30412 31202: contig of 791 bp in length
*      31203 31302: gap of 100 bp
*      31303      32082: contig of 780 bp in length
*      32083 32182: gap of 100 bp
*      32183 32934: contig of 752 bp in length
*      32935 33034: gap of 100 bp
*      33035 33796: contig of 762 bp in length
*      33797 33896: gap of 100 bp
*      33897      34683: contig of 787 bp in length
*      34684 34783: gap of 100 bp
*      34784 35552: contig of 769 bp in length
*      35553 35652: gap of 100 bp
*      35653 36406: contig of 754 bp in length
*      36407 36506: gap of 100 bp
*      36507      37294: contig of 788 bp in length
*      37295 37394: gap of 100 bp
*      37395 38159: contig of 765 bp in length
*      38160 38259: gap of 100 bp
*      38260 39017: contig of 758 bp in length
*      39018 39117: gap of 100 bp
*      39118      39890: contig of 773 bp in length
*      39891 39990: gap of 100 bp
*      39991      40774: contig of 784 bp in length
*      40775 40874: gap of 100 bp
*      40875 41643: contig of 769 bp in length
*      41644 41743: gap of 100 bp
*      41744      42521: contig of 778 bp in length
*      42522 42621: gap of 100 bp
*      42622 43393: contig of 772 bp in length
*      43394 43493: gap of 100 bp
*      43494 44268: contig of 775 bp in length
*      44269 44368: gap of 100 bp
*      44369      45149: contig of 781 bp in length
*      45150 45249: gap of 100 bp
*      45250 46024: contig of 775 bp in length
*      46025 46124: gap of 100 bp
*      46125 46868: contig of 744 bp in length
*      46869 46968: gap of 100 bp
*      46969      47720: contig of 752 bp in length
*      47721 47820: gap of 100 bp
*      47821 48596: contig of 776 bp in length
*      48597 48696: gap of 100 bp
*      48697 49462: contig of 766 bp in length
*      49463 49562: gap of 100 bp
*      49563      50335: contig of 773 bp in length
*      50336 50435: gap of 100 bp
*      50436 51207: contig of 772 bp in length
*      51208 51307: gap of 100 bp
*      51308 52078: contig of 771 bp in length
*      52079 52178: gap of 100 bp
*      52179 52954: contig of 776 bp in length
*      52955 53054: gap of 100 bp
*      53055      53813: contig of 759 bp in length

```



50914	CAAAGAGGAGCTCTCGGAAAATGAGGATGAGCAGAACAGCCACCAAAAA	50865			
1287	ysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLys	1303			
50864	AGGGCAAAAGAGGAGGAGCCACCAAAACCTCTTGGTGGGGGACATCGAAG	50815			
1304	GluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerG1	1320			
50814	GAAGAGCCCAATGAAACATCCAAAGAAGGAAACAAAGAAAAAATCTGT	50765			
1320	yProProAlaProGluGluGluGluGluGluGluGluGlnSerGlyAsnT	1337			
50764	ACCTCCTGTAGTACGACGATGAAGAAGAGAAACAAATTTGGAACA	50715			
1337	hrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGln	1353			
50714	CAGACATAAGTCAAAAGACAAACAGCAGCCGCACATCAAAGAGACACA	50665			
1354	.Gln. ....	1354			
50664	CAGAGGTGAGTGTGTGCAGTGTCTAAAGTGGAGCTTCGGTGCAGTAAATGTG	50615			
1354	.....	1354			
50614	CTTACCAATAATGTAATCATATTTGATGCTATCCACATTTGGGTCTTCC	50565			
1355	.....ArgAlaGluSerProGluSerAlaIleGluSerThrGlnS	1369			
50564	CCAAGCAGCAGCAGAAATCTCTGAAACAAGTGCAGTTGAATCCACACAGT	50515			
1369	erThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer	1385			
50514	CCACACCACAGAAAGCAGGAGAGACCATCAAAAGCACCATCACCATCA	50465			
1386	GlnProLysLysAsnVal	1391			
50464	CAACCCCAAAAAAATG	50447			
seq_name: gb_htg5:AC016449					
seq_documentation_block:					
LOCUS	AC016449	72157 bp	DNA	HTG	13-
DEFINITION	Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING				
ACCESSION	AC016449				
VERSION	AC016449.2 GI:9104517				
KEYWORDS	HTG; HTGS_PHASE0.				
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 72157)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
TITLE	Homo sapiens, clone RP11-45L14				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 72157)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckle,A., Castle,A., Colangelo,M., Collins,S., Collymore,C., Dearellano,K., Dewar,K., Domino,M., Donelan,F., Ferraire,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,R., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatsis,L., Lechoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melton,D., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Tait, Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Jul 13, 2000 this sequence version replaced qi:6471.				

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced qi:6479175.



```

1355 .....ArgAlaGluSerProGluSerSe 1362
29032 TATCCACATTGGTCTTCCCAACGACGAGATCTCCTGAATCTAG 28983
1362 rAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProS 1379
28982 TGCAATTGAATCCACACATCCACACACAGAAAGGACGAGGAAGACCAT 28933
1379 erlYsThrProSerProSerGlnProLysLysAsnVal 1391
28932 CAAAACGCCATCACCATCACAAACCAAAAAAATGTG 28895
seq_name: gp_htg22:AL512630

seq_documentation_block:
LOCUS   AL512630 340969 bp  DNA  HTG  21-MAR-2001
DEFINITION Mus musculus chromosome 5 clone RP21-583E8, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL512630
VERSION   AL512630.2 GI:13443466
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 340969)
AUTHORS   Sims, S.
TITLE     Direct Submission
JOURNAL   Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT   On Mar 24, 2001 this sequence version replaced gi:12193249.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: dm583E8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 323281 bases at least Q40
Consensus quality: 329435 bases at least Q30
Consensus quality: 333234 bases at least Q20
Insert size: 336469; sum-of-contigs
Insert size: 189900; 1.3% error; agarose-fp
Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
coverage: 11.68x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source           1..340969
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP21-583E8"
                     /clone_lib="RPCI-21"
     misc_feature     1..10854
                     /note="assembly_fragment:06116
                     fragment_chain:1"
     misc_feature     10955..16884
                     /note="assembly_fragment:05022
                     fragment_chain:1"
     misc_feature     16985..22344
                     /note="assembly_fragment:04089
                     fragment_chain:1"
     misc_feature     22445..36631
                     /note="assembly_fragment:00629
                     fragment_chain:1"
                     /note="assembly_fragment:04629
                     fragment_chain:1"
                     /note="assembly_fragment:02431
                     fragment_chain:1"
                     /note="assembly_fragment:06652
                     fragment_chain:2"
                     /note="assembly_fragment:02856
                     fragment_chain:2"
                     /note="assembly_fragment:05282
                     fragment_chain:2"
                     /note="assembly_fragment:01199
                     fragment_chain:2"
                     /note="assembly_fragment:00854
                     fragment_chain:3"
                     /note="assembly_fragment:04769
                     fragment_chain:3"
                     /note="assembly_fragment:03592
                     fragment_chain:3"
                     /note="assembly_fragment:02797
                     fragment_chain:4"
                     /note="assembly_fragment:05253
                     fragment_chain:4"
                     /note="assembly_fragment:01663
                     fragment_chain:4"
                     /note="assembly_fragment:03653
                     fragment_chain:5"
                     /note="assembly_fragment:05568
                     fragment_chain:5"
                     /note="assembly_fragment:01977
                     fragment_chain:5"
                     /note="assembly_fragment:00023
                     fragment_chain:6"
                     /note="assembly_fragment:05254
                     fragment_chain:6"
                     /note="assembly_fragment:00245
                     fragment_chain:7"
                     /note="assembly_fragment:03587
                     fragment_chain:7"
                     /note="assembly_fragment:00642
                     fragment_chain:8"
                     /note="assembly_fragment:04656
                     fragment_chain:8"
                     /note="assembly_fragment:01666
                     fragment_chain:9"
                     /note="assembly_fragment:05529
                     fragment_chain:9"
                     /note="assembly_fragment:02430
                     fragment_chain:10"
                     /note="assembly_fragment:00629
                     fragment_chain:10"

```



```
* 15233 15912: contig of 680 bp in length
* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp
* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length
```

```
* 44741 44840: gap of 100 bp
* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length
* 49510 49609: gap of 100 bp
* 49610 50331: contig of 722 bp in length
* 50332 50431: gap of 100 bp
* 50432 51135: contig of 704 bp in length
* 51136 51235: gap of 100 bp
* 51236 51981: contig of 746 bp in length
* 51982 52081: gap of 100 bp
* 52082 52783: contig of 702 bp in length
* 52784 52883: gap of 100 bp
* 52884 53580: contig of 697 bp in length
* 53581 53680: gap of 100 bp
* 53681 54398: contig of 718 bp in length.
FEATURES
    source
        Location/Qualifiers
            1. 54398
                /organism="Homo sapiens"
alignment_scores:
    Quality: 819.50      Length: 196
    Ratio: 5.028        Gaps: 1
    Percent Similarity: 83.163    Percent Identity: 83.163
alignment_block:
    US-09-512-581-2 x AC068224/rev
Align seg 1/1 to reverse of: AC068224 from: 1 to: 54398
1227 GluLysLeuGlyMetAspLeuThrLysLeuValGlnGluLysPr 1243
29482 GAGAAATTAGGTATGATGACTGAACCTAAGTTGGTACAGGAACAACCC 29433
1243 OlyGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerA 1260
29432 TAAAGGCAGTCAGCGAAGTCGGAANAGAGGCCATACGGCTTCAGAATCTG 29383
1260 spGluGlnGlnTrpProGluGluLysArgLysGluAspLeuGlu 1276
29382 ATGACACAGCTGGCCTCAGGAAAAGAGGCTCAAAGAGATATATTAGAA 29333
1277 AsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPr 1293
29332 AATGAAGATGAACAGACATAGTCCGCCCAAAAAGGCTAAAAGAGCGGACC 29283
1293 oProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysT 1310
29282 ACCAAAACCTCTTGTGTGAGGTACACCAAAAAGAGAGCCAAACAATGAAA 29233
1310 hrSerLysLysGlySerLysLysLysSerGlyProProAlaProGluGlu 1326
29232 CTTCTAAAAGAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGGAG 29183
1327 GluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSe 1343
29182 GAGGAAGAAGAAAGAAAGACAAAGTGGAAATACGGAACAGAGAGTCCAAAAG 29133
1343 rLysGlnHisArgValSerArgArgAlaGlnGln..... 1354
29132 CAACAGACCCGAGTGTCAAGGAGAGACAGCAGAGAGTAAAGCATGTGTAA 29083
1354 ..... 1354
29082 CTCTAAACTGCATCTGTTTCGTTACTATATTATAAATCATATTGTATGC 29033
```



```

:::||||:..... |||||| |
5333 AAAATATTCAGCTGAAATCCACAGAACTTTGGATGGAGTTTGACTACGTA 5382
1083 r.....SerLeuGluSerProLysAspProValLeuProAlaArpPhep 1098
5383 TCCAAAGCAAAATTTAAATACCTTACGAAATTTTAAAGCAAAATTCCTCTA 5432
1098 heThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
5433 TAGATGAGAAAACAGAAATTTTCAAT.....AAATCTTCATACGCGCT 5476
1115 GluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValle 1131
5477 AAAATGGAAAGT.....CAGATTGA 5496
1131 uGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL 1148
5497 ACATGCTATTAAGACACACAGCTCATCTTTTGGGAACAGACTACCAATA 5546
1148 ysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
5547 AGCAGCAAAATTTGAGAAAAGAACTCATAGCTCAAGACCGATATAA 5596
1165 SerSerProGlyArgLysGlyArgLeuAspSerSerGluMetAspHl 1181
5597 AAATCGTCTAGAGAGGAAA.....AACGAAAAAAGGAG 5631
1181 sSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLys 1198
5632 AAAGCTAAATGACGACCAATCTTAATATTAGGACGTCGCCAAGACGTA 5681
1198 erAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPro 1214
5682 GTCAGAGATTCAGGGATTCGTATTAATATTTCGGAAGCACCGCTCT... 5728
1215 ArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMe 1231
5729 .....TCAAGTGAGGAGATAAGCGAAGAGAGAGGAGGAAATTAGCGA 5769
1231 tAspAspLeuThrLysLeu 1237
5770 AGAAGATTTTGTATGAATA 5788
seq_name: gb_htg14:AC068224

```

```

seq_documentation_block:
LOCUS AC068224 54398 bp DNA HTG 30-APR-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC068224
VERSION AC068224.1 GI:7671284
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
Boguslavsky,L., Boukhgatter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kan,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
Boguslavsky,L., Boukhgatter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kan,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

```

Meidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

# TITLE JOURNAL COMMENT

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: LI0161  
Center clone name: 660\_H\_19  
-----

\* NOTE: This record contains 68 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

```

* 1 685: contig of 685 bp in length
* 686 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp
* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 12885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp

```







[illegible]







1355 .....ArgAlaGluSerProGluS 1361  
 21608 ATGCTATCCACATTTGGGTCTTCCCAAGCAGACAGACTCTCTGAAT 21657  
 1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArg 1377  
 21658 CTAGTGAATTAATTCACACAGTCCACACCACAGAAAGGACGAGGAAGA 21707  
 1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391  
 21708 CCATCAAAACGCCATCACCATCAACCAACCAAAAAAATGTG 21749

seq\_name: gb\_pr5:AL138820

seq\_documentation\_block:  
 LOCUS AL138820 168487 bp DNA PRI 01-NOV-2000  
 DEFINITION Human DNA sequence from clone RP11-380B4 on chromosome 13, complete sequence.

ACCESSION AL138820  
 VERSION AL138820.11 GI:11022364  
 KEYWORDS HTG.  
 SOURCE human.

#### ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

AUTHORS Smith,M.

TITLE Direct Submission

JOURNAL Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonequery@sanger.ac.uk

On Oct 25, 2000 this sequence version replaced gi:10715762.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13  
 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

RP11-380B4 is from the library RPC1-11.2 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see http://bacpac.med.buffalo.edu/

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-380B4.

#### FEATURES

##### source

1. .168487  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="RP11-380B4"  
 /clone\_lib="RPC1-11.2"

BASE COUNT 48433 a 33155 c 34624 g 52275 t

##### ORIGIN

#### alignment\_scores:

Quality: 925.50

Ratio: 5.057

Length: 214

Gaps: 1

Percent Similarity: 85.514 Percent Identity: 85.514

#### alignment\_block:

US-09-512-581-2 x AL138820

Align seg 1/1 to: AL138820 from: 1 to: 168487

1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluG1 1225  
 |||||  
 20594 TCTGAATTTGAGAAAGCCTAGAGGAGGAAAAAAGCCCGTCACAGACA 20643  
 |||||  
 1225 nGluGluLysLeuGlyMetAspLeuThrLysLeuValGlnGluGlnL 1242  
 |||||  
 20644 GGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGAACAGA 20693  
 |||||  
 1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258  
 |||||  
 20694 AACCTAAGGCAGTCAGCGAAGTCGAAAGAGAGGCCATACGGCTTCAGAA 20743  
 |||||  
 1259 SerAspGluGlnGlnTTPProGluGluLysArgLeuLysGluAspIleLe 1275  
 |||||  
 20744 TCTGATGAACAGCAGTGGCTGAGGAAAAAGAGGCTCAAGAAGATATATT 20793  
 |||||  
 1275 uGluAsnGluAspGluGlnAsnSerProLysLysGlyLysArgGlyA 1292  
 |||||  
 20794 AGAAATGAAGATCAACAGAAATAGTCGCCCAAAAGGGTAAAGAGGCC 20843  
 |||||  
 1292 rgProProLysProLeuGlyGlyThrProLysGluGluProThrMet 1308  
 |||||  
 20844 GACCACCAAAACCTCTCTGGTGGAGGTACACCAAAAGAGAGGCCAACATG 20893  
 |||||  
 1309 LysThrSerLysLysGlySerLysLysSerGlyProProAlaProG1 1325  
 |||||  
 20894 AAAACTTCTAAAGAAAGGAAACAAATAATCTGGACCTCCAGCACCAGA 20943  
 |||||  
 1325 uGluGluGluGluGluGlnArgGlnSerGlyAsnThrGluGlnLysSerL 1342  
 |||||  
 20944 GGAGGAGGAGAAAGCAAGAAACACAAAGTGGAAATACGGAACAGAGTCCA 20993  
 |||||  
 1342 ysSerLysGlnHisArgValSerArgArgAlaGlnGln..... 1354  
 |||||  
 20994 AAAGCAACAGCACCAGTGTCTCAAGGAGAGCAGACAGAGTAAGCATGT 21043  
 |||||  
 1354 ..... 1354  
 21044 GTAACCTAAACTGCATCTGTTTGTCTACTATATATAATCATANTTTG 21093  
 |||||  
 1355 .....ArgAlaGluSerProGluS 1361  
 |||||  
 21094 ATGCTATCCACATTTGGGTCTTCCCAAGCAGACAGACTCTCTGAAT 21143  
 |||||  
 1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArg 1377  
 |||||  
 21144 CTAGTGAATTAATTCACACAGTCCACACCACAGAAAGGACGAGGAAGA 21193  
 |||||  
 1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391  
 |||||  
 21194 CCATCAAAACGCCATCACCATCAACCAACCAAAAAAATGTG 21235  
 |||||

seq\_name: gb\_pl4:SPAC

seq\_documentation\_block:

LOCUS SPAC 9982 bp DNA PLN 16-SEP-2000  
 DEFINITION S.pombe chromosome 1 cosmid cl10.

ACCESSION AL441624

VERSION AL441624.1 GI:10185123

KEYWORDS 5'-amp-activated protein kinase; cdc42; cell division control  
 protein 42 homolog; chromosome dynamics; heat shock protein 70-like  
 protein; pss1; sp1.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;







```

|||||
388 TCTGATGAACAGCGTCCCTGAGGAAAGAGGCTCAAGAAGATATTT 437
1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA 1292
438 AGAAATGAAGATGAACAGAAATAGTCGCCCAAAAGGGTAAAGAGGCC 487
1292 rgProProLysProLeuGlyGlyThrProLysGlyGluGluProThrMet 1308
488 GACCACCAAAACCTCTGGTGGAGGTACACCAAAAGAGAGACCAACATG 537
1309 LysThrSerLysLysGlySerLysLysLysSerGlyProProLysProGln 1325
538 AAAACTTCTAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGA 587
1325 uGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerL 1342
588 GGAGGGAAGGAAGGAAGAAAGACAAAGTGGAAATACGGAACAGAGTCCA 637
1342 ySerLysGlnHisArgValSerArgArgAlaGlnGln..... 1354
638 AAGCAACAGCAGCCGAGTGTCAAGGAGACACAGCAGAGAGTAAAGCATGT 687
1354 ..... 1354
688 GTAACCTCTAAATGCATCTGTTTGGTTACTATATATATAATCATATTTG 737
1355 ..... ArgAlaGluSerProGluS 1361
738 ATGCTATCCACATTTGGGTCTCTCCCAAGCAGACAGAGATCTCCTGAAT 787
1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArg 1377
788 CTAGTGCATTTGAATCCACAGTCCACACACACAGCAAAAGGACGAGGA 837
1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
838 CCATCAAAACGCCATCACCATCACACCAACCAAAAAAATGTG 879

seq_name: gb_pr8:HS49J10

seq_documentation_block:
LOCUS HS49J10 137246 bp DNA PRI 22-NOV-1999
DEFINITION Human DNA sequence from PAC 49J10, BRCA2 gene region chromosome
13q12-13 contains ESTs.
ACCESSION Z84572
VERSION Z84572.1 GI:1813974
KEYWORDS 13q12-13.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137246)
Direct Submission
Whiteley, M.
Submitted (09-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT:
This sequence is not the entire insert of clone . It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 49J10 is at 1 in this sequence. The true
right end of clone 267P19 is at 43908.
The true left end of clone 179I15 is at 137143.
49J10 is from the human PAC library described in Ioannou A.P. et al

```

---

```

FEATURES
            source
            Location/Qualifiers
                1..137246
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="13"
                /map="13q12-13"
                /clone="XX-49J10"
                1551..1831
                /note="AluX repeat: matches 281. .1 of consensus;
                incomplete repeat"
                repeat_region
                1838..1871
                /note="17 copies of 2 mer 100 % conserved"
                repeat_region
                1882..2741
                /note="L1PA15 repeat: matches 893. .1 of consensus"
                repeat_region
                2597..2848
                /note="L1 repeat: matches 5389. .5139 of consensus"
                repeat_region
                3149..3405
                /note="MER44A repeat: matches 247. .1 of consensus"
                repeat_region
                3853..3922
                /note="AluSc repeat: matches 188. .120 of consensus;
                incomplete repeat"
                repeat_region
                5620..5810
                /note="MIR repeat: matches 58. .262 of consensus"
                repeat_region
                7829..8133
                /note="AluX repeat: matches 1. .301 of consensus"
                repeat_region
                10957..11242
                /note="AluY repeat: matches 291. .1 of consensus"
                repeat_region
                12344..12405
                /note="L1PA16 repeat: matches 903. .842 of consensus"
                repeat_region
                12381..13119
                /note="L1PA16 repeat: matches 833. .68 of consensus"
                repeat_region
                14428..14540
                /note="HY1 repeat: matches 110. .1 of consensus"
                repeat_region
                14951..15071
                /note="MIR2 repeat: matches 126. .1 of consensus"
                repeat_region
                16589..16885
                /note="AluX repeat: matches 302. .4 of consensus"
                repeat_region
                18067..18397
                /note="L1 repeat: matches 3807. .4127 of consensus"
                repeat_region
                19076..19481
                /note="MLT1A2 repeat: matches 374. .1 of consensus"
                repeat_region
                19516..19799
                /note="AluX repeat: matches 19. .302 of consensus;
                incomplete repeat"
                repeat_region
                19808..20127
                /note="AluY repeat: matches 2. .299 of consensus"
                prim_transcript
                <21359..26899
                /note="match: multiple ESTs; match: H44781 D20082 W20070
                T64867 H45693; match: F12354 N91467 N83765 H45741 H93424"
                repeat_region
                24795..25089
                /note="AluX repeat: matches 294. .1 of consensus"
                repeat_region
                26083..26106
                /note="12 copies of 2 mer 96 % conserved"
                prim_transcript
                complement(27120..27476)
                /note="match: multiple ESTs; match F09972 T72607"
                prim_transcript
                <27633..28101
                /note="match: multiple ESTs; match: 5' EST T51695 clone
                72405; Paired with EST T51543 matching this clone; match
                H91918"
                repeat_region
                28327..28360
                /note="17 copies of 2 mer 88 % conserved"
                prim_transcript
                <28864..>29013
                /note="match: 3' EST T51543 clone 72405; Paired with EST
                T51695 matching this clone"
                repeat_region
                29300..29456
                /note="MIR2 repeat: matches 146. .1 of consensus"
                repeat_region
                30648..30934
                /note="AluX repeat: matches 7. .302 of consensus"
                repeat_region
                33156..33439
                /note="AluX repeat: matches 301. .2 of consensus"
                repeat_region
                34900..35062
                /note="L1MD2 repeat: matches 425. .263 of consensus"

```







60843 TCAGGCCCAAGCAGCCCAACATTCGCCAAGTGGGTCTATAATAAGATTTTC 60799

525 .....IlePheSerLysValMetValIleThr.ArgAsnLeuP 537  
::: ||| ||||::: ::: ||| ||||  
60793 TTTTGATAGTTGTAACACAGTCATGTTTTTTTCTTCAGACTGCCTC 60744

537 roAspProGLyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu 553  
|||||| | |||| ::::::: ::::: |||::: |||  
60743 CAGATCCCCTAAAGCGCAGAGTACCCTAACCCAGTTTAGCAACAACCTG 60694

554 GluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProTh 570  
::: ||| ::::::: ::::: ||||::: ||||:::  
60693 CGAAAAGATGTCAGCTCCTACGATGCATTAAACATGTCCTTAAACCGCA 60644

570 rCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysL 587  
|||||| | ||||::: ||::: ::::: |||||  
60643 CGTAAGCTGTGGGAGTGTGCTGACACAGATGGGAGTCTCTCTGAAMAAC 60594

587 euGlyAsnProLysGlnProThrAsnPropHeLeuGluMetIleLysPhe 603  
||||| ||| ::::: ||||::: ::::: ||||:::  
60593 TTGGCGCCACAGTCCCAA...TCGAATCTGTATTACACACAGTTAAGATG 60547

604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAl 620  
|||||| | ||||::: ||| ::::: |||||  
60546 CTGATGTGACGCGTGGCATCAGTCATGTTGGACACAGGAGTCTATTGGCGT 60497

620 aleuIleLys.....GlnValAsnLysSerIleAspGlyThrAlaAsp 623  
:|||||::: ::::: ||||::: ||||:::  
60496 ACTTATATAGTAAATTGCTTTTAAGTGGCCTTTTGCCAGACTTAATCCGG 60447

624 .....GlnValAsnLysSerIleAspGlyThrAlaAsp 634  
::: ||||::: ||||::: ||||:::  
60446 TAATCTTTTGTCTCAGCCTAATTGAACAATGCATCGAAGGGTCCATG 60397

635 AspGluAspGluGlyValProThrAspGlnAlalleArgAlaGlyLeuGl 651  
||||| ::::::: ||||::: ||||::: ||||:::  
60396 TGTGAAGAGATCGGAATTTTCGGCTCAGGAAGCGGCGAGCGGACTCAA 60347

651 uLeuLeu..Lys.....PheLeuSerPheThrHisProIleSerP 654  
:|||||::: ::::: ||||::: ||||:::  
60346 GTTGTGGCCGTGAGTATAGAGATAATTTAAGTTATGATACAGAATCCA 60297

655 .....ValLeuSerPheThrHisProIleSerP 664  
::: ||||::: ||||::: ||||:::  
60296 TTTGATGCTTTATCTTGCAGATGCTTTTCATAGCTCTTTCCGGCTCAT 60247

664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680  
||||| ::::::: ||||::: ||||::: ||||:::  
60246 TCTTCACCGACTCTCGCTGGCTCATTTAAATTTCCCTGCTCAGTTACGAG 60197

681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly.. 696  
::: ||||| ::::::: ||||::: ||||::: ||||:::  
60196 CAGGATTATGTTGCACCACTGGTGTAAAGACATAACTCATCTGGGGCG 60147

697 .....SerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711  
|||||| | ||||::: ||| ||||::: ||||  
60146 CTATCAGCCACTCATTCAGACCCCTACACCGGCTATTCTCCAGNGCTGG 60097

711 euProVal.....LeuHisLysSerLysLysGlyProProArg 724  
||||| ::::::: ||||::: ||||::: ||||:::  
60096 CTCGGGTGTGAAGGACTTCCACTT.....ATTGGAACCCCCAAG 60057

725 GlnAlaLysTyrrAlaIleHisCystIleHis..... 734  
|||||| | ||||::: ||||::: ||||:::  
60056 CANGCAAAAGCACGGGTGCGATGATATTTTCGTTAAACGCCAGTCTCGCGC 60007

735 .....AlaIlePheSerSerLysGluThrG 743  
||| ||||  
60006 TTCCTACTGATGGAGCAACAAGTGGAGCTGGAAGTGCCTCCCACTACAC 59957

743 In.....PheAlaGlnIlePheGluProLeuHisLysSer 754  
|| ||| ::::::: ||||::: ||||::: ||||:::  
59956 AAACAGTGCATCTATTTTCAACAGAGATCATCTCGACAGCGCTTCGCTAA 59907

[illegible]



[illegible]

61741 ..ATTACAAACAAAATCTATGATATCATTTTTCAGACTTAACCGGCATCAAT 61694  
267 SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLys 283  
SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLys 283  
61693 GCAGCGCTGTGGTGTTTCAGTGTAGTCTCAGCTGAGGAACAACCTGCTGTC 61644  
283 rAsnAspAsnGluGluArgLeu..... 290  
61643 CACGGACCATGCTGAGAGCTAAGTAGCTAATTTATCAGCATGTACCTCT 61594  
291 .....GlnValValLysLeu 295  
61593 ATTTCGATATTATAAACATATTTTTTCGTACTACAGAGCCAACTCTC 61544  
296 LeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLy 312  
61543 TTGTCCTCGATGTTCTCGAAAGAGACTCGCAACTAGCGAAAAAGTACC 61494  
312 sProLeuTrpGlnCysTyrrLeuGlyArgPheAsnAspIleHisValProI 329  
61493 CAACCTGTTGAAATCTTCTTGGCGCTTCTGGCACATCACCGAACCA 61444  
329 leArgLeuGluCysValLysPheAlaSerHisCysLysLeuMetAsnHisPro 345  
61443 TTCGCATCAAGTGTTCTCAGTCATCCATCGCACTTCTCTGCTCAATCAC 61394  
346 AspLeuAlaLysAspLeuThrGluTyrrLeuLysValArgSerHisAspPr 362  
61393 AGCTCTCAGCAGCATATCACCGAANAATTCGGCTTCGAAATCAGGATCT 61344  
362 ogLuGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL 379  
61343 TCACGAAGTGGTCCGCCACGAGGTGGTAATGGCTATTCTGGAACCTGCCA 61294  
379 ysLysAspIleLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal 393  
61293 AGCGGACTTCACCTCGTCTCGAAGCGCCGATCTACTAGAATTTGTG 61244  
394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410  
61243 CGCGAGCGGACGCTAGATAAGAAGTACAAAATTCGACGGATGCGATCAA 61194  
410 tGlyLeuAlaGlnIleTyrrLysTyrrAlaLeuGlnSerAlaAlaGlyL 427  
61193 TGGTCTGGCGCTACATCTACAAGCGCGCAATTTGCGAACCCAACTCTAA 61144  
427 ysAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis 442  
61143 GCACGTGGCCTTAAGCTCAGGTGTAGTGTAGTAAGAACAAGTACTAGT 61094  
443 IleTyrrGlnAnSerIleAspAspArgLeuValGluArgIlePh 459  
61093 GGATACTACAAAGTGGGCTTGGAGGACCGCTGCTGTGTGGAACGCGCTACT 61044  
459 eAlaGlnTyrrMetValProHisAsnLeuGluThrThruArgMetLysC 476  
61043 TATCAGCTCGCTGGTTCCTATAAATAGTCCGGAAGCGCATGAAGA 60994  
476 ysLeuTyrrTyrrLeuAlaThrLeuAspLeuAlaValLysAlaLeu 492  
60993 ACCTATATCAITTTGCTAGCGCATCTCGATGCCAATGCCAACAGGCTTC 60944  
493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509  
60943 GTCGAGCTGCAAGAACAACAAATGAAGACACAGCAATACGGTCAGTGTG 60894  
509 uLeuAspLeuIleLysGln.....ProLysThrAspAlaSerV 522  
60893 GATCAAGCTACACCACTCTAAGGAGTTTCCGCGCGGCTACTTGTAGTCAGC 60844  
522 alLysAla..... 524



36289	AAGACCTGCCGTTCGGAATTTCCAAAACAATGAGCTCTATATACCGCTG	36333
1115	GluMetLysSerPheThrProGlyLysProLysThrThrAsnValLe	1131
36339	GAGGTG.....TATACGCTGGGAGCCAAATCCACG.....	36368
1131	uGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL	1148
36369	.....AGCAAAGCTG	36378
1148	ysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsnPro	1164
36379	CGCAACAGCAATGACACGTCGGCGAGCAGCAGTGGCTCCAAAGAGACCG	36428
1165	Ser.....Ser.ProGlyArgIleLysGlyArgLeuAspSerSerGluM	1179
36429	CGGGAACAGTCAATCATGGATGATGATAAATCCGACGGTGTGAGGTGCT	36478
1179	eAspHisSerGluAsnGluasp.....	1186
36479	TGGAATAACTGAGAGAATGAATCTAACTAGTGTGTAATTTGATTTCAGG	36528
1187	.....TyThrMetSerSerProLe	1193
36529	AGAACAATCTGTCGACAAACATAGACGGCGTGACACTACGAGGCCCATG	36578
1193	uProGlyLysLysSerAspLys	1200
36579	GCCAAACGAACGCGCGCAGGAG	36600
seq_name: gb_htg6.AC020286		
seq_documentation_block:		
LOCUS	AC020286	194634 bp DNA HTG
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS * pieces.	
ACCESSION	AC020286	
VERSION	AC020286.1	GI:6664611
KEYWORDS	HTG; HTGS_PHASE2.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Pterygota; Neoptera; Endopterygota; Diptera; Brachyces Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
AUTHORS	1 (bases 1 to 194634)	
TITLE	Adams,M. and Venter,J.C.	
JOURNAL	Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Rockville, MD, USA	
COMMENT	This sequence was identified as CDM:10212663 by the s For more information on this record e-mail to fly@cel * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available * the accession number will be preserved.	
FEATURES	Location/Qualifiers	
source	1..194634	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
BASE COUNT	55777 a 42412 c 41301 g 55144 t	
ORIGIN		
alignment_scores:		
Quality:	1708.50	Length: 1498
Ratio:	1.975	Gaps: 46
Percent Similarity:	57.744	Percent Identity: 30.975
alignment_block:		
US-09-512-581-2 x AC020286/rev ..		
Align seq 1/1 to reverse of: AC020286 from: 1 to: 194634		











Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster* Science 287 (5461), 2185-2195 (2000)

20196006

2 (bases 1 to 262731)

Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

On Oct 9, 2000 this sequence version replaced gi:7303489.

Location/Qualifiers

## source

1. .262731

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/chromosome="2R"

complement(join(5117, 5856, 6026, .6362, 6424, .7856, 8828, .9717, 17398, .17559, 17629, .17763, 22517, .23464))

/gene="otk"

/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."

/product="CT25769"

/db\_xref="FLYBASE:FBan0008967"

/db\_xref="FLYBASE:FBgn0004839"

complement(<5117, .>23464)

/gene="otk"

/note="otk"

/map="48D8-48E2"

/db\_xref="FLYBASE:FBan0008967"

/db\_xref="FLYBASE:FBgn0004839"

complement(join(5764, .5856, 6026, .6362, 6424, .7856, 8828, .9717, 17398, .17559, 17629, .17763, 22517, .22568))

/gene="otk"

/note="otk gene product: Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008967"

/db\_xref="FLYBASE:FBgn0004839"

/protein\_id="AAF58596.1"

/db\_xref="GI:7303541"

/translation="WTARMISICGLVVALMWASVLASSRRQRVPQSQSVVENESVKE ECESTDSTSEUHWDLNHRHAYDKRVHQIGSNLHIEAVRRTEDVGNVCIATNLAS GARQAPAKLUSVLYLESASVOLGSSNRNELLKCHVGESGDLEPLEIEWRNSEKL STWRNVDQHLRIIROPGSEDGLYRCTASNAAGVMSKQGYVQSVKCLPRLPRLL KNEKMSWDKOTIFLCRGKGAAGLEALPAAPEDRLVQGPIGQSIKEGEPTALTC LYEPLDELKNRIOLRWKDKGLLRQVELGSGAPICGHSFDGSKDALLREDARLVLHK QNGTLSFASIIASDAGYQCOLLEAHNPINSFGLIEVIEQLKEVPQPTSKNLELDA VVAKVHCQAQGTPTFPVQWRDGNTEPLPDHVEVDANGTLIFRNWNSHRGNITCLAT NSQOINATVATINVPVQFPVPGTSETQGVVHMQAIGDPKPTIOWDKLKY LSENNDRERFRLENGTLEIRNVQVEDESGYCTIGNSAGLKREDVQLVVKTGDDG APESGGDGLVLTFTVVALAYIVLVVGLMWCYRQARKARLNDLSTKEAGG DQPDVAGKSGSEQPCLSKOHNGHSKSRKSGDAOKSDPTACSQOSRASKSAHIY BOLALPRSGLSELTOIGRGEFGDVFKLKLATLVTSFSDKADATEKHNSNSGSGS GSGSTLTNLEKRKRSMDDIIEIEEEDQDQHNQSGLEQLVLVVKALNKVKDEQAQO

## mRNA

complement(join(5117, 5856, 6026, .6362, 6424, .7856, 8828, .9717, 17398, .17559, 17629, .17763, 22517, .23464))

## gene

complement(join(5117, 5856, 6026, .6362, 6424, .7856, 8828, .9717, 17398, .17559, 17629, .17763, 22517, .22568))

/gene="otk"

/note="otk gene product: Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008967"

/db\_xref="FLYBASE:FBgn0004839"

/protein\_id="AAF58596.1"

/db\_xref="GI:7303541"

/translation="WTARMISICGLVVALMWASVLASSRRQRVPQSQSVVENESVKE ECESTDSTSEUHWDLNHRHAYDKRVHQIGSNLHIEAVRRTEDVGNVCIATNLAS GARQAPAKLUSVLYLESASVOLGSSNRNELLKCHVGESGDLEPLEIEWRNSEKL STWRNVDQHLRIIROPGSEDGLYRCTASNAAGVMSKQGYVQSVKCLPRLPRLL KNEKMSWDKOTIFLCRGKGAAGLEALPAAPEDRLVQGPIGQSIKEGEPTALTC LYEPLDELKNRIOLRWKDKGLLRQVELGSGAPICGHSFDGSKDALLREDARLVLHK QNGTLSFASIIASDAGYQCOLLEAHNPINSFGLIEVIEQLKEVPQPTSKNLELDA VVAKVHCQAQGTPTFPVQWRDGNTEPLPDHVEVDANGTLIFRNWNSHRGNITCLAT NSQOINATVATINVPVQFPVPGTSETQGVVHMQAIGDPKPTIOWDKLKY LSENNDRERFRLENGTLEIRNVQVEDESGYCTIGNSAGLKREDVQLVVKTGDDG APESGGDGLVLTFTVVALAYIVLVVGLMWCYRQARKARLNDLSTKEAGG DQPDVAGKSGSEQPCLSKOHNGHSKSRKSGDAOKSDPTACSQOSRASKSAHIY BOLALPRSGLSELTOIGRGEFGDVFKLKLATLVTSFSDKADATEKHNSNSGSGS GSGSTLTNLEKRKRSMDDIIEIEEEDQDQHNQSGLEQLVLVVKALNKVKDEQAQO

## CDS

complement(join(5117, 5856, 6026, .6362, 6424, .7856, 8828, .9717, 17398, .17559, 17629, .17763, 22517, .22568))

/gene="otk"

/note="otk gene product: Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008967"

/db\_xref="FLYBASE:FBgn0004839"

/protein\_id="AAF58596.1"

/db\_xref="GI:7303541"

/translation="WTARMISICGLVVALMWASVLASSRRQRVPQSQSVVENESVKE ECESTDSTSEUHWDLNHRHAYDKRVHQIGSNLHIEAVRRTEDVGNVCIATNLAS GARQAPAKLUSVLYLESASVOLGSSNRNELLKCHVGESGDLEPLEIEWRNSEKL STWRNVDQHLRIIROPGSEDGLYRCTASNAAGVMSKQGYVQSVKCLPRLPRLL KNEKMSWDKOTIFLCRGKGAAGLEALPAAPEDRLVQGPIGQSIKEGEPTALTC LYEPLDELKNRIOLRWKDKGLLRQVELGSGAPICGHSFDGSKDALLREDARLVLHK QNGTLSFASIIASDAGYQCOLLEAHNPINSFGLIEVIEQLKEVPQPTSKNLELDA VVAKVHCQAQGTPTFPVQWRDGNTEPLPDHVEVDANGTLIFRNWNSHRGNITCLAT NSQOINATVATINVPVQFPVPGTSETQGVVHMQAIGDPKPTIOWDKLKY LSENNDRERFRLENGTLEIRNVQVEDESGYCTIGNSAGLKREDVQLVVKTGDDG APESGGDGLVLTFTVVALAYIVLVVGLMWCYRQARKARLNDLSTKEAGG DQPDVAGKSGSEQPCLSKOHNGHSKSRKSGDAOKSDPTACSQOSRASKSAHIY BOLALPRSGLSELTOIGRGEFGDVFKLKLATLVTSFSDKADATEKHNSNSGSGS GSGSTLTNLEKRKRSMDDIIEIEEEDQDQHNQSGLEQLVLVVKALNKVKDEQAQO

## mRNA

complement(join(<27067, .27801, 27890, .28383, 28455, .>28527))

/gene="CG8964"

/product="CT25760"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(<27067, .>28527)

/gene="CG8964"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(join(27067, .27801, 27890, .28383, 28455, .28527))

/gene="CG8964"

/note="CG8964 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

/protein\_id="AAF58595.1"

/db\_xref="GI:7303540"

/translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR EDSRVLKSGSGLQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQOP TSKNLELGTLSKVHCKAQGTAPQVKKMRETQLPLPNVTDQNGTLIFNQVSNQRQGY TCIAINSQGOITATVINSVVAAPKFSVPPPEGEVIAEAGTAVIHCOAIGPKPTIQW KDULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYCTIGSAGLKREDVLLVLK SSKASNSIVTRIIIVICLAFVLVLGLKVMVRYRRLKVKQLEDGHVNGPTDGO EHDHNEPCLTEANSKSLKRESTILBQESQVADDIV"

complement(join(29253, .29670, 29731, .30211, 30362, .30624, 30681, .30790, 30862, .31313))

/gene="CG8889"

/product="CT25522"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(<29253, .>31313)

/gene="CG8889"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(join(29483, .29670, 29731, .30211, 30362, .30624, 30681, .30720))

/gene="CG8889"

/note="CG8889 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

/protein\_id="AAF58594.1"

/db\_xref="GI:7303539"

/translation="MAQSSWQPIDCKDNCTRLLLIADPQILGNSYDRSSHSPLARYD SDRYLAKTFERALFTQPHITVFLGDLDEGNIATQAYKQYQVFRRIYQNKYKKR VHVPGDNDIGGENDYISNSQRRFENFEDLFDYDNLRFFKINMLDLSNPDR DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIFSCWHESRIIFYPSTKVINFY ENSVRFPDLKALKEHSYLEIMVPTCSYRMKSKIGIGYAVLENLYSLTVLQPNR FILLFTYVFWGLFVVCGFVFPKMTCPFRVAKRQTLNRYSTIPOP"

join(31631, .31871, 32192, .32294, 32356, .32741, 32805, .33012, 33071, .33239, 33303, .34044, 34109, .34372, 34430, .34523, 34586, .35431, 35497, .35633, 35695, .35774, 35836, .36101, 36174, .36464, 36528, .>36620)

/gene="CG17509"

/product="CT32429"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

<31631, .>36620

/gene="CG17509"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

## gene

complement(join(<27067, .27801, 27890, .28383, 28455, .>28527))

/gene="CG8964"

/product="CT25760"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(<27067, .>28527)

/gene="CG8964"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(join(27067, .27801, 27890, .28383, 28455, .28527))

/gene="CG8964"

/note="CG8964 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

/protein\_id="AAF58595.1"

/db\_xref="GI:7303540"

/translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR EDSRVLKSGSGLQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQOP TSKNLELGTLSKVHCKAQGTAPQVKKMRETQLPLPNVTDQNGTLIFNQVSNQRQGY TCIAINSQGOITATVINSVVAAPKFSVPPPEGEVIAEAGTAVIHCOAIGPKPTIQW KDULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYCTIGSAGLKREDVLLVLK SSKASNSIVTRIIIVICLAFVLVLGLKVMVRYRRLKVKQLEDGHVNGPTDGO EHDHNEPCLTEANSKSLKRESTILBQESQVADDIV"

complement(join(29253, .29670, 29731, .30211, 30362, .30624, 30681, .30790, 30862, .31313))

/gene="CG8889"

/product="CT25522"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(<29253, .>31313)

/gene="CG8889"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(join(29483, .29670, 29731, .30211, 30362, .30624, 30681, .30720))

/gene="CG8889"

/note="CG8889 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

/protein\_id="AAF58594.1"

/db\_xref="GI:7303539"

/translation="MAQSSWQPIDCKDNCTRLLLIADPQILGNSYDRSSHSPLARYD SDRYLAKTFERALFTQPHITVFLGDLDEGNIATQAYKQYQVFRRIYQNKYKKR VHVPGDNDIGGENDYISNSQRRFENFEDLFDYDNLRFFKINMLDLSNPDR DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIFSCWHESRIIFYPSTKVINFY ENSVRFPDLKALKEHSYLEIMVPTCSYRMKSKIGIGYAVLENLYSLTVLQPNR FILLFTYVFWGLFVVCGFVFPKMTCPFRVAKRQTLNRYSTIPOP"

join(31631, .31871, 32192, .32294, 32356, .32741, 32805, .33012, 33071, .33239, 33303, .34044, 34109, .34372, 34430, .34523, 34586, .35431, 35497, .35633, 35695, .35774, 35836, .36101, 36174, .36464, 36528, .>36620)

/gene="CG17509"

/product="CT32429"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

<31631, .>36620

/gene="CG17509"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

## CDS

complement(join(<27067, .27801, 27890, .28383, 28455, .>28527))

/gene="CG8964"

/product="CT25760"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(<27067, .>28527)

/gene="CG8964"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(join(27067, .27801, 27890, .28383, 28455, .28527))

/gene="CG8964"

/note="CG8964 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

/protein\_id="AAF58595.1"

/db\_xref="GI:7303540"

/translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR EDSRVLKSGSGLQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQOP TSKNLELGTLSKVHCKAQGTAPQVKKMRETQLPLPNVTDQNGTLIFNQVSNQRQGY TCIAINSQGOITATVINSVVAAPKFSVPPPEGEVIAEAGTAVIHCOAIGPKPTIQW KDULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYCTIGSAGLKREDVLLVLK SSKASNSIVTRIIIVICLAFVLVLGLKVMVRYRRLKVKQLEDGHVNGPTDGO EHDHNEPCLTEANSKSLKRESTILBQESQVADDIV"

complement(join(29253, .29670, 29731, .30211, 30362, .30624, 30681, .30790, 30862, .31313))

/gene="CG8889"

/product="CT25522"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(<29253, .>31313)

/gene="CG8889"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(join(29483, .29670, 29731, .30211, 30362, .30624, 30681, .30720))

/gene="CG8889"

/note="CG8889 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

/protein\_id="AAF58594.1"

/db\_xref="GI:7303539"

/translation="MAQSSWQPIDCKDNCTRLLLIADPQILGNSYDRSSHSPLARYD SDRYLAKTFERALFTQPHITVFLGDLDEGNIATQAYKQYQVFRRIYQNKYKKR VHVPGDNDIGGENDYISNSQRRFENFEDLFDYDNLRFFKINMLDLSNPDR DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIFSCWHESRIIFYPSTKVINFY ENSVRFPDLKALKEHSYLEIMVPTCSYRMKSKIGIGYAVLENLYSLTVLQPNR FILLFTYVFWGLFVVCGFVFPKMTCPFRVAKRQTLNRYSTIPOP"

join(31631, .31871, 32192, .32294, 32356, .32741, 32805, .33012, 33071, .33239, 33303, .34044, 34109, .34372, 34430, .34523, 34586, .35431, 35497, .35633, 35695, .35774, 35836, .36101, 36174, .36464, 36528, .>36620)

/gene="CG17509"

/product="CT32429"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

<31631, .>36620

/gene="CG17509"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

## mRNA

complement(join(<27067, .27801, 27890, .28383, 28455, .>28527))

/gene="CG8964"

/product="CT25760"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(<27067, .>28527)

/gene="CG8964"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(join(27067, .27801, 27890, .28383, 28455, .28527))

/gene="CG8964"

/note="CG8964 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

/protein\_id="AAF58595.1"

/db\_xref="GI:7303540"

/translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR EDSRVLKSGSGLQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQOP TSKNLELGTLSKVHCKAQGTAPQVKKMRETQLPLPNVTDQNGTLIFNQVSNQRQGY TCIAINSQGOITATVINSVVAAPKFSVPPPEGEVIAEAGTAVIHCOAIGPKPTIQW KDULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYCTIGSAGLKREDVLLVLK SSKASNSIVTRIIIVICLAFVLVLGLKVMVRYRRLKVKQLEDGHVNGPTDGO EHDHNEPCLTEANSKSLKRESTILBQESQVADDIV"

complement(join(29253, .29670, 29731, .30211, 30362, .30624, 30681, .30790, 30862, .31313))

/gene="CG8889"

/product="CT25522"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(<29253, .>31313)

/gene="CG8889"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(join(29483, .29670, 29731, .30211, 30362, .30624, 30681, .30720))

/gene="CG8889"

/note="CG8889 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

/protein\_id="AAF58594.1"

/db\_xref="GI:7303539"

/translation="MAQSSWQPIDCKDNCTRLLLIADPQILGNSYDRSSHSPLARYD SDRYLAKTFERALFTQPHITVFLGDLDEGNIATQAYKQYQVFRRIYQNKYKKR VHVPGDNDIGGENDYISNSQRRFENFEDLFDYDNLRFFKINMLDLSNPDR DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIFSCWHESRIIFYPSTKVINFY ENSVRFPDLKALKEHSYLEIMVPTCSYRMKSKIGIGYAVLENLYSLTVLQPNR FILLFTYVFWGLFVVCGFVFPKMTCPFRVAKRQTLNRYSTIPOP"

join(31631, .31871, 32192, .32294, 32356, .32741, 32805, .33012, 33071, .33239, 33303, .34044, 34109, .34372, 34430, .34523, 34586, .35431, 35497, .35633, 35695, .35774, 35836, .36101, 36174, .36464, 36528, .>36620)

/gene="CG17509"

/product="CT32429"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

<31631, .>36620

/gene="CG17509"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

## gene

complement(join(<27067, .27801, 27890, .28383, 28455, .>28527))

/gene="CG8964"

/product="CT25760"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(<27067, .>28527)

/gene="CG8964"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

complement(join(27067, .27801, 27890, .28383, 28455, .28527))

/gene="CG8964"

/note="CG8964 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008964"

/db\_xref="FLYBASE:FBgn00033674"

/evidence=not\_experimental

/protein\_id="AAF58595.1"

/db\_xref="GI:7303540"

/translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR EDSRVLKSGSGLQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQOP TSKNLELGTLSKVHCKAQGTAPQVKKMRETQLPLPNVTDQNGTLIFNQVSNQRQGY TCIAINSQGOITATVINSVVAAPKFSVPPPEGEVIAEAGTAVIHCOAIGPKPTIQW KDULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYCTIGSAGLKREDVLLVLK SSKASNSIVTRIIIVICLAFVLVLGLKVMVRYRRLKVKQLEDGHVNGPTDGO EHDHNEPCLTEANSKSLKRESTILBQESQVADDIV"

complement(join(29253, .29670, 29731, .30211, 30362, .30624, 30681, .30790, 30862, .31313))

/gene="CG8889"

/product="CT25522"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(<29253, .>31313)

/gene="CG8889"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

complement(join(29483, .29670, 29731, .30211, 30362, .30624, 30681, .30720))

/gene="CG8889"

/note="CG8889 gene product"

/codon\_start=1

/db\_xref="FLYBASE:FBan0008889"

/db\_xref="FLYBASE:FBgn00033675"

/protein\_id="AAF58594.1"

/db\_xref="GI:7303539"

/translation="MAQSSWQPIDCKDNCTRLLLIADPQILGNSYDRSSHSPLARYD SDRYLAKTFERALFTQPHITVFLGDLDEGNIATQAYKQYQVFRRIYQNKYKKR VHVPGDNDIGGENDYISNSQRRFENFEDLFDYDNLRFFKINMLDLSNPDR DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIFSCWHESRIIFYPSTKVINFY ENSVRFPDLKALKEHSYLEIMVPTCSYRMKSKIGIGYAVLENLYSLTVLQPNR FILLFTYVFWGLFVVCGFVFPKMTCPFRVAKRQTLNRYSTIPOP"

join(31631, .31871, 32192, .32294, 32356, .32741, 32805, .33012, 33071, .33239, 33303, .34044, 34109, .34372, 34430, .34523, 34586, .35431, 35497, .35633, 35695, .35774, 35836, .36101, 36174, .36464, 36528, .>36620)

/gene="CG17509"

/product="CT32429"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"

<31631, .>36620

/gene="CG17509"

/map="48E1-48E2"

/db\_xref="FLYBASE:FBan0017509"

/db\_xref="FLYBASE:FBgn00033676"







```

147549  ....  |||...:|||||  |||...:|||||...:|||||  147598
AGCTTTACACGATATCACCGAGAAATTGCGGCTTCGAAATCACGACTCT  147598

362  oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL  379
...:|||||...:|||||...:|||||...:|||||...:|||||
TGACGAAGTGGTCCGCCACGAGGTGTAATGCTATTGTGGAACAGTCCA  147648

379  yLysAspIleLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal  393
|||...:|||||...:|||||...:|||||...:|||||...:|||||
AGCGGACTTCACCCCTCGTTCTCGAAGCGCCGCTACTAGAAATTGTG  147698

394  ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe  410
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CGGAGCGGAGCTAGATAAGAGTACAAATTCGACGGGTGCGATGAA  147748

410  tGluLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL  427
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TGGTCTGGCTACATCTACAGCGCGCAATTTCGGAACCCACGATCTAA  147798

427  yAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis  442
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GCACGTGGCTTAAGGTCAAGGTGACTGGATTAAGAACAGATACTACAT  147848

443  IleTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIlePh  459
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GGATACTACAAAGTGGCTTGGAGGACCGCTTGTGGACGCGCTACT  147898

459  eAlaGlnTyrMetValProHisAsnLeuLeuThrThrGluArgMetLysC  476
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TATCACCTCCCTGGTTCCTATATACTAGCTCCGGAGAGCGCATGAAGA  147948

476  yLeuTyrTyrLeuTyrAlaThrLeuAspLeuAlaValLysAlaLeu  492
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
AGCTATATCATTTGTAGCGATCTCGATGCCAATGCCACCAAGGCTTC  147998

493  AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe  509
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GTGAGCTGCAGAAGAACCAATGAAGACACGCAATACCGTCAGTATTG  148048

509  uLeuAspLeuIleLysGln.....ProLysThrAspAlaSerV  522
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GATCAAGCTACACCACTCTCAAGGAGTTCACGCGCGCTACTAGTCAGC  148098

522  allysala.....  524
...:|||||

148099  TCAGGCCCAAGCAGGCGCAACATTGCCAAGTGGTCTATAATAAGTTTC  148148

525  .....IlePheSerLysValMetValIleThr.ArgAsnLeup  537
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TTTTGTAGTTTGTACCAAGTATGTTGTTTTTTTCTTCAGACTGCTTC  148198

537  roAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu  553
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CAGATCCCTTAAAGCGGAGAGTACCTAACCCAGTTTAGCAACAACCTG  148248

554  GluAspAspGluLysIleArgLysGlnLeuLeuValValSerProTh  570
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CGAAAGATGCTCAGCTCCTCAGTGCATTAACATTTGCTTTAAACGCGA  148298

570  rCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLysL  587
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CGTAAGCTGCGGAGTGTGTCACACGATGGAGTTCTCTGAAAAACG  148348

587  euGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPhe  603
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TTGGCGCCACAGTCCAA...TCGAATCTGATTACAACACAGTAAAGATG  148395

604  LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAl  620
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CTGATTGACGCGTGGCATCAGTGATGGTGGACAAAGGAGTCTATTGGCGT  148445

620  aLeuIleLys.....  623
:|||||...:

```

```

148446  ACTTATTAGGTAAATTGCTTTTAAGTGGCCCTTTTGCCAGACTTAATCCGG  148495

624  .....GlnValAsnLysSerIleAspGlyThrAlaAsp  634
...:|||||...:|||||...:|||||...:|||||...:|||||
TAATCTTTTGTCTCAGCCTTAATTGAACAATGCATCGAAGAGGGTCCATG  148496

635  AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGl  651
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TGTGAAGAGATCGGAATTTCGGCTCAGGAAGCGGCGGAGGACTCAA  148546

651  uLeuLeu...Lys.....  654
:|||||...:

148596  GTTGTGGCCGTGAGTATAGAGATAATTTAAGTTATGATACAGAAATCCA  148645

655  .....ValLeuSerPheThrHisProIleSerP  664
...:|||||...:|||||...:|||||...:|||||...:|||||
TWTGATTGCTTTATATACTTCAGAGATGCTTTTCATAGCTTTTCGGCTCAT  148646

664  heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp  680
|| ...:|||||...:|||||...:|||||...:|||||...:|||||
TCTTCACCGATACTTCGGCTGCGTCAATTTAATTCCTGCTCAGTTACGAG  148696

681  AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly...  696
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CAGGATTATGTCACCACTGGTGTAAAGACACTAACTCATCTGGGGCG  148746

697  .....SerLysIleGluLysPheProHisIleArgSerAlaLeuL  711
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CTATCAGCCACTCATTCAGACCCCTACACCGGCTATTCTCGACGAGCTGG  148796

711  euProValLeuHisLysSerLysLysGlyProProArgGlnAlaLys  727
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CTCCGGTGTGCAAGGACTTCGCACTATTGGAAACCCCAAGCAAGCAAG  148846

728  TyrAlaIleHisCysIleHis.....  734
:|||||...:

148896  CACGGTGCAGTATTTTCTTAACACCCAGTCGTGCGCTTCCACTGA  148945

735  .....AlaIlePheSerSerLysGluThrGln.....  743
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TGGAGCAACAAGTGGAGCTGGAAGTGCCTCCCACTACACAACAGTCG  148946

744  .....PheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro  757
||| ...:|||||...:|||||...:|||||...:|||||...:|||||
ATCCCTATTTTCAACGAGATCATCGAGACCTTCGCGCTAAACACTGACAC  148996

758  SerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAl  774
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
...ACTGTGACATCAGCGCACAAAGATTGTGACCTTGGGTCACTATGC  149046

774  aLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlar  791
| ...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
CTACAATATGCCACAGGCTTCTTAACGCCCATTAAGAATATGATTGCGC  149093

791  hrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys  807
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GACGAATTGTCAAAGAGCTGTATCCCAAGAA...GTTCCTGCGCAGCGA  149143

808  ThrThrLysLeu.....TrpValProAspGluGluValSe  819
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GACTAGCACTGCCAGACGACAGTGTGCTGTCGCCAAGAGAAACTACC  149190

819  rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL  836
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
GCCGACACTCTATGTAAGCTGGATCGCTCAAGGCTATGGCCAGGTGGC  149240

836  euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg  852
|||||...:|||||...:|||||...:|||||...:|||||...:|||||
TTTTGGGATTGCTACCGGAT...GAGCAGCTGCCCCAGAAAACTCCGA  149290

853  LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLy  869
...:|||||...:|||||...:|||||...:|||||...:|||||...:|||||
ATGCTGGCGCGCTTGTGTTAAACCAACGAGCGGATTGCTCGGCCAGACCG  149337

```







```
869 sIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaI 886
      :::::  |||  |||||  |||||  :::::  |||
25110  WTTCGCGTGCAGAAATCTGGCTCGCGCTCGGGCAGCTGCGCCA 25159
      |||||  |||||  |||||  |||||  |||||  |||||
886  leValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlu 902
      :::::  |||  |||||  |||||  :::::  |||
25160  TGCTCAAGTGTGGACAAAAGGCGTAGGTGATCAGTACAGCGCTGAG 25209
      |||||  |||||  |||||  |||||  |||||  |||||
903  GlnTyr...GlnLeuCysAlaLeuAlaIleAsnAspGlu..... 914
      |||||  |||||  |||||  |||||  |||||  |||||
25210  CAGTATTCAGCTTCCAGCTGATGGTGGTAGAGACATTCATCTTGT 25259
      |||||  |||||  |||||  |||||  |||||  |||||
914  ..... 914
25260  GGTAGAACTTAGTCTCAATGCTATCTCGGTGTTCCCGTAGGTGATCC 25309
      |||||  |||||  |||||  |||||  |||||  |||||
915  ..CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeu 930
      |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  |||||
25310  GGTGCCA.GAAGTTCGGGAATCTTGTCTCGCAAGCTGCACAAAGGATTA 25358
      |||||  |||||  |||||  |||||  |||||  |||||
931  SerArgLeuArg.....LeuProLeuGluTyrMetAlaIleCys 943
      |||||  |||||  |||||  |||||  |||||  |||||
25359  AGCAGAAGTTTCCAGGAAGTGTGTCGCGCTGGACTTCATGGCTTGA 25408
      |||||  |||||  |||||  |||||  |||||  |||||
943  salalaLeuCysAla.....LysAspProValLysGluArgA 955
      ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
25409  TGTGTCGCTGGCTAGAGACTGAGGAGAGTGCAGTAGTTCAAGAAACCA 25458
      |||||  |||||  |||||  |||||  |||||  |||||
955  rgAlaHisAlaArg.GlnCysLeuValLys..... 964
      ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
25459  TCGGATCCGTCGGAGAGTATGTGATAAATATTTCAACAGATTGCAAGA 25508
      |||||  |||||  |||||  |||||  |||||  |||||
965  .....AsnIleAsnValArgArgGluTyrL 973
      ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
25509  CCTTGTGCTCACTATGCAGAACGATGATAAACAAACGGCGGAATATC 25558
      |||||  |||||  |||||  |||||  |||||  |||||
973  euLys..... 974
      |||||  |||||  |||||  |||||  |||||  |||||
25559  TCAAGACTGTGCTATGACATGATAGTGTGTTTACTTTTGGAAATCATATG 25608
      |||||  |||||  |||||  |||||  |||||  |||||
975  .....GlnHis.AlaAlaValS 980
      ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
25609  TGTATTTTAAACCATTAATCAATATTATCCAGCTCCCGACAGCTCAACGG 25658
      |||||  |||||  |||||  |||||  |||||  |||||
980  erGluLysLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
      ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
25659  AGTCACAATCATTAACATACACTACCTGACTACATGCTGCTTCGCTATT 25708
      |||||  |||||  |||||  |||||  |||||  |||||
997  HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluI 1013
      |||||  |||||  |||||  |||||  |||||  |||||
25709  CCGGTGCTGGTCCAGATCCAGCTTCAGAAATCAGAGGACTGATACA 25758
      |||||  |||||  |||||  |||||  |||||  |||||
1013  nLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetA 1030
      |||||  |||||  |||||  |||||  |||||  |||||
25759  GCTGGCAGATGGAGAGTGGCTGGCTTTCATTCTGGAGCGCTGATGG 25808
      |||||  |||||  |||||  |||||  |||||  |||||
1030  laLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
      |||||  |||||  |||||  |||||  |||||  |||||
25809  CCAAAAGCAGAAACGTTGTGCCATAGCTTCTACAAGCAGCTGCTGCAGCTG 25858
      |||||  |||||  |||||  |||||  |||||  |||||
1047  IleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn... 1062
      |||||  |||||  |||||  |||||  |||||  |||||
25859  ATAAAGCATCGGAGTTACGTCTGGGG...TCGGACAAGCGCGCAACATTA 25905
      |||||  |||||  |||||  |||||  |||||  |||||
1062  ..... 1062
25906  TGTATCTCACTCAATCATTTTCCAGGCTTTCCCTTCTAAATATATTTA 25955
      |||||  |||||  |||||  |||||  |||||  |||||
1063  .....GluLysLeuTyrThrValCysAspValAla 1072
      |||||  |||||  |||||  |||||  |||||  |||||
25956  TACATATATTTTCGGTTATCAGAAATGTGGCGCTGCTGCGATCTTGGC 26005
      |||||  |||||  |||||  |||||  |||||  |||||
1073  MetAsnIleIleMetSerLys.....SerThrTh 1082
```

```
|||  |||||  |||||  :::::  |||
26006  ATGTACATTATCGACTCCAAAGTTTCAGTCCATTGTGCAACACGAGCAC 26055
      |||||  |||||  |||||  |||||  |||||  |||||
1082  rTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheT 1099
      |||||  |||||  |||||  |||||  |||||  |||||
26056  CTTTCCATGCGCGTG.....GCTTGCAGAAATGATTATATA 26093
      |||||  |||||  |||||  |||||  |||||  |||||
1099  hrGlnProAsp...LysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
      |||||  |||||  |||||  |||||  |||||  |||||
26094  AAGAGCTCGCGCTTGCGAATTTCCAAACAATGACGTCTATATACCGCTG 26143
      :::::  |||  |||||  |||||  :::::  |||
1115  GluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValle 1131
      ::|||  ::|||  |||||  |||||  |||||  |||||
26144  GACGTG.....TATACGTGGGAGCAAAATCCACG..... 26173
      |||||  |||||  |||||  |||||  |||||  |||||
1131  uclyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL 1148
      |||||  |||||  |||||  |||||  |||||  |||||
26174  .....ACCAAAGCTG 26183
      |||||  |||||  |||||  |||||  |||||  |||||
1148  ysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
      ::|||  ::|||  |||||  |||||  ::|||  ::|||
26184  CGCAACAGCAATGACACGTCGCGAGCAGCAGTGGCTCCAAAGAGACGG 26233
      |||||  |||||  |||||  |||||  |||||  |||||
1165  Ser.....Ser.ProGlyArgIleLysClyArgLeuAspSerSerGluM 1179
      ::|||  ::|||  |||||  |||||  ::|||  ::|||
26234  CGGGAACAGTCAATCATGATGATGAAATCCGACGAGTGAGTGCTCT 26283
      |||||  |||||  |||||  |||||  |||||  |||||
1179  etAspHisSerGluAsnGluAsp..... 1186
      ::|||  ::|||  |||||  |||||  ::|||  ::|||
26284  TGGAAATACTGAGATGAATCTAAGTCTGCTGTAATTTGATTTCAGG 26333
      |||||  |||||  |||||  |||||  |||||  |||||
1187  .....TyrThrMetSerSerProle 1193
      |||||  |||||  |||||  |||||  |||||  |||||
26334  AGAACAACTGTTTCGACAACTACGAGCGCTGACACTACGAGGCCCATG 26383
      |||||  |||||  |||||  |||||  |||||  |||||
1193  uproGlyLysLysSerAspLys 1200
      |||||  |||||  |||||  |||||  |||||  |||||
26384  GCCAACGAACCGCGCAGGAG 26405
      |||||  |||||  |||||  |||||  |||||  |||||
seq_name: gb_inl.AC007474
seq_documentation_block:
LOCUS AC007474 192763 bp DNA INV 28-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
BACR14L19, complete sequence.
ACCESSION AC007474
VERSION AC007474.5 GI:13162477
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 192763)
Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Busan,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ikegami,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirska,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 49A-49B
Unpublished
2 (bases 1 to 192763)
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.W., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
```



23272 AGTCTTCAGCAGCATATCACCGAGAAATTCGGGCTCCGAAATCACGATCT 23321  
 362 oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL 379  
 23322 TGACGAAGTGGTCCGCCACGAGGTGGTAATGGCTATTGTGAAACATGCCA 23371  
 379 yLysAspIleLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal 393  
 23372 AGCGGACTTCACCCCTCGTCTCGAAGCGCCGATCTACTAGAAATGTG 23421  
 394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410  
 23422 CGCGAGCGGCGGTAGATAGAAAGTCAAAAATTCGACGGATCGATCAA 23471  
 410 tGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427  
 23472 TGGTCTGGCCCTACATCTACAAAGCGCGCAATTCGGAACCCCAACGATCAA 23521  
 427 yAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis 442  
 23522 GCACCTGGCCTTAAGGTCAGGTTGACTGGATTAAACAACAGATACAT 23571  
 443 IleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePh 459  
 23572 GGATACCTACAAAGTGGCTTGGAGGACCGCTGCTTGGAAACGCTACT 23621  
 459 eAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysC 476  
 23622 TATCACCTGGCTGGTCCCTTAACTAGTTCGGGAAGCGCATGAAGA 23671  
 476 yLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeu 492  
 23672 AGCTATATCATTTGCTAGGCGATCTCGATGCCAATGCCACCAAGCTTTC 23721  
 493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509  
 23722 GTCAGCTGCAGAGAAACAAATGAAGACACGCAATACGGTTCAGTGATG 23771  
 509 uLeuAspLeuIleLysGln.....ProLysThrAspAlaSerV 522  
 23772 GATCAAGCTACACCACTCTAAGGAGTTTCAGCGCGCGTACTTAGTCAGC 23821  
 522 alLysAla.....IlePheSerLysValMetValIleThr..ArgAsnLeup 524  
 23822 TCACGCCCAAGCAGGCCAACATTCCCAAGTGGTCTCTATAATAGTTTC 23871  
 525 .....IlePheSerLysValMetValIleThr..ArgAsnLeup 537  
 23872 TTTTGATAGTTTGTAAACAGTGTGTTTTTTTTTTCTTCAGACTGCTTC 23921  
 537 roAspProGlyLysAlaGluAspPheMetLysLysPheThrGlnValLeu 553  
 23922 CAGATCCCTTAAAGCGGAGTACCTAACCCAGTTTAGCAACACCTG 23971  
 554 GluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProTh 570  
 23972 CGAAAAAGATGCTCAGCTCCACGATGATTAACATTGCTCTTAACCGCA 24021  
 570 rCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysL 587  
 24022 CGTAAGCTGTGGGAGTGTGTCACACGATGGGAGTTCTCTGAAAAAGC 24071  
 587 euGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603  
 24072 TTGGCCGCCACGTCCAA...TCGAATCTGTATTACAAACACAGTTTAAGATG 24118  
 604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAl 620  
 24119 CTGATTGAGCGCGTGGCATCAGTGATGGTGACACAGGAGTCTATTGGCGT 24168  
 620 aLeuIleLys..... 623  
 24169 ACTTATTAGTAAATTCGTTTTAAGTGCCTTTTTCGACAGACTTAATCCGG 24218

624 .....GlnValAsnLysSerIleAspGlyThrAlaAsp 634  
 24219 TAATCTTTTTCAGCCTAATTGAACAATGCATCGAGAAAGGTCCATG 24268  
 635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuG1 651  
 24269 TGTGAAGAGATCGAAATTCGGCTCAGGAAGCGGCGGCGGACTCAA 24318  
 651 uLeuLeu...Lys..... 654  
 24319 GTTGTGGCGCTGAGTAGAGATAATTTAAGTTATGATACAGAAATCCA 24368  
 655 .....ValLeuSerPheThrHisProIleSerP 664  
 24369 TTTGATTGCTTTATACCTGCGATGCTTTTCATACGCTCTTTTCGGCTCAT 24418  
 664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680  
 24419 TCITCACCGATACCTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAG 24468  
 681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly... 696  
 24469 CAGGATTATGTCACCACTGCTGCTAAAGACACTAACTCATCTGGGCGC 24518  
 697 .....SerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711  
 24519 CTATCACCCACTCATTCAGCACCCCTACCGGCTATTCTCGACGAGCTGG 24568  
 711 euProValLeuHisHisLysSerLysGlyProProArgGlnAlaLys 727  
 24569 CTCGGGTGCAAGGACTTCGCACCTTATTGGAACCCCAACGAAGCAAG 24618  
 728 TyrAlaIleHisCysIleHis..... 734  
 24619 CACGCGTGGGATGATTATTCGTAACAGCAGCTCGTCGGCTTCCACTGA 24668  
 735 .....AlaIlePheSerLysGluThrGln..... 743  
 24669 TGGAGCAACAAGTGGAGCTGGAGTGGCTCCACAACACTACACAACAGTGC 24718  
 744 .....PheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757  
 24719 ATCTATTTTCAACGAGATCATCGAGAGCTTCGCCTTAAACTGACACCA 24768  
 758 SerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAl 774  
 24769 ...AACTGTGAGCATCAGCGCACAAAGATTGTGACCTTGGGTCACTGC 24815  
 774 aLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAla 791  
 24816 CTACATATGCCACAGGCTTCTTAACGCCCATTAAGAATATGATGCGC 24865  
 791 hrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807  
 24866 GACGAATTGTCAAAGAGCTGCTTATCCAAGAA...GTTCTCGCCGACGA 24912  
 808 ThrThrLysLeu.....TrpValProAspGluGluValSe 819  
 24913 GACTACGAACCTGCCAGAGCAGTGTGTTGGTCCCAAGAGAAATACC 24962  
 819 rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836  
 24963 CGCGAGACCTCTATGTAGCTGGATGCGCTCAAGGCTATGCCAGGTGGC 25012  
 836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852  
 25013 TTTTGGGATTCGATACCGAT...GAGCAGCTGCGCCAGAAACATTCGA 25059  
 853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLy 869  
 25060 ATGCTGCGCGCTTGTGTTAAACCAACGAGGCGATTGTGTCGCCAAGACCG 25109







[illegible]

2190	GTAAGATAAGGAA.....ATTAAATCTGATCAGGCTACC.....	2225	
1238	lGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisT	1255	1255
2226	.....CAGGCCACATCAGCAGTGCACGAGGAAAGAAACAGTAA	2268	2268
1255	hrAlaSerGluSerAspGlu...GlnGlnTrpProGluGluLysArgLeu	1270	1270
2269	CAGCAGCTGGTCAGAGAATATCAACAACAAACACAGATGAGAA.....	2312	2312
1271	LysGluAspIleLeuGlnGluAsnGluAspGluGlnAsnSerPro...ProLys	1286	1286
2313	.....GTAGATGAATCGGGACCTCCCGCCCTTC	2341	2341
1286	sLysGlyLysArgGlyArgProProLysProLeuGly...GlyGlyThrP	1302	1302
2342	CAAAACCCAGAGAGAGCTGCACCAAGTCTGAATCTCAGGGCAATGCTA	2391	2391
1302	roLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys	1318	1318
2392	CCAAAAATGATGATCTAAATAACATATTAAACAAGGGAAGGAGAGCT	2441	2441
1319	..SerGlyProAlaPro.....Glu	1325	1325
2442	GCAGTGGGTGAGGAGAGCCCTGGGGTTTGAAGCAGGTAATGCCAAAGC	2491	2491
1326	GluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys.....	1340	1340
2492	ACCCAACTGCAAGATTGACCAAAAGGAGCAGCAGCAGGAAAGACAAA	2541	2541
1341	.....SerLysSerLysGlnHis.....ArgValSerArgAlaG	1353	1353
2542	TTGACTTACAAAGGTAAATAATGATTTGCAAGGGAGAAATGAAGGCCA	2591	2591
1353	lGlnArgAlaGluSerProGluSerSer	1362	1362
2592	AAACGAAG...CAGGCTCCAGCTTCTGCA	2617	2617
seq_name: gb_inl:AC007475			
seq_documentation_block:			
LOCUS	AC007475	173613 bp	DNA
DEFINITION	Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone BACR04E2R, complete sequence.		
ACCESSION	AC007475		
KEYWORDS	AC007475.7	GI:13162478	
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacle,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.		
TITLE	Sequencing of Drosophila chromosome 2R, region 49A-49B		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 173613)		
AUTHORS	Celisner,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhof,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,		



```

gene      1. .2558
/ gene="KIAA0648"
CDS       <1. .2558
/ gene="KIAA0648"
/ codon_start=3
/ product="KIAA0648 protein"
/ protein_id="BA331623.1"
/ db_xref="GI:3327110"
/translat="L Y L Y L A S Q D P N A V K A L N E M W K C C M L R S H V R E L L D L H K Q T S E A
N C S A M F G L M T I A K N L P D P G K A D F V K F N Q V L G D E K L R S Q L E L L I S P T C S C G K Q A D I
C V R E I A R K L A N P O P T N P F L E M V K F T L P A H I D E A I S A L V K L M K N S I E G T A D D E
E E G V S P D T A I R S G L E L L K L V S F T H P T S F S A E T Y E S L L O C L R M E D D K V A E A A I O I P N
T G H I E R D L P Q I R S T L I P I L H O K A K R G T P H O A Q A V H C I H A I F T N K E V Q L A Q I F E P L S
R S N A D V E Q L I T P L V S I G H I S M L A P D F A S P M K S V V A N F I V K D L L M N D R S T G E K N K
L W S P D E V S E V L A K V A I K L L V R M L G M K N Q S K S A N S T L R L S A M L V S E G D I T E O K
R I K S D M S R L A A G S A I M K L A O P C V Y H E I T P E Q F O L C A L V I N D E C Y Q V R I F A Q K L
H K A L V K L L P L E Y M A I F A L C A K D P V K E R R A H R O C L K N I S I R R E Y I K O N P M A T E K L L
S L L P E Y V P Y M I H L A D H P D T R S O D V D L R D I K E C L F M L E V L M T K N E N N S H A F M K K
M A E N I K L T R A Q S P D E S K T N E K L Y C D V A L C V I N S K A L C N A D S P K D P V L P M K F F T Q
P E K D F C N D K S Y I S E E T R V L L T G K P G A V G A V N K P L S A T G R K P Y V R S T G T E T G S N I
N V S E L N P S T G N R S R E Q S E A A E T G V S E N E E N P V R I I S V T P V K N I D P V K N K E I N S D Q A
T Q G N I S D R G K R T V T A A G A E N I Q O K T D E K V D E S G P P A P S K P R R G R P R K S E S O G N A T K
N D D L N K I N G R R A A V Q E S P G L E A G N A K P K L Q D L A K A A P A E R Q I D L Q R"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN

alignment_scores:
Quality: 2704.50 Length: 910
Ratio: 3.675 Gaps: 15
Percent Similarity: 80.879 Percent Identity: 60.549

alignment_block:
US-09-512-581-2 x AB014548 ..
Align seg 1/1 to: AB014548 from: 1 to: 5177

477 LeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAs 493
|||||
3 TTATATTACTTATATGCTAGTTGGATCCAAATGCTGTAAAAAGCTCTCAA 52

493 nGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuL 510
|||||
53 CGAATGTGGAAGTGTACAGACATGCTTCGGAGCCCATGTACCGCAACTAT 102

510 euAspLeuLeuLysGlnProLysThrAspAlaSerValLysAlaLeuPhe 526
|||||
103 TGGATTTCACACAGCAGCCTACATCAGAGGCTAACTGTCTGCCATGTTT 152

527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGl 543
|||||
153 GGAAACTGTATGACCATAGCAAGAAATTTGCCCTGACCCCGGGAAAGCACA 202

543 nAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleA 560
|||||
203 AGATTTTGAAGAAATTTAAACAGGTTCTCGCGCATGATGAGAAACATTC 252

560 rGlyGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
|||||
253 GCTCTCAGTTGGAGTATTATTAATAGCCCAACCTGTTCTTTGCAACAAGCA 302

577 GluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPr 593
|||
303 GATATTGTGTGAGAGAAATACCCCGGAACCTTGCAAAATCCTTAAGCAACC 352

593 ofThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaP 610
|||||
353 ACAAAATCTTTTCTAGATAGGTCAAAATTTCTGTGGAAAGATCGCAC 402

610 rovalHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
|||||
403 CTGTGCACATTGATTCAGAGGCAATAGTGCACTAGTGAATGATGAAT 452

```

```

627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAs 643
|||||
453 AAGTCAATAGAGGGACAGCAGATGATGAAGAGGGGTGAAGTCCACA 502

643 pGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThr 660
|||||
503 TACAGCTATCCGTTCCAGGACTTGAACCTTCTTAAGGTTCTGTCTTTACAC 552

660 IspProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
|||||
553 ATCTTACCTCGTTCCACTCTGCAGAGACATATAGTCTTGTACAGTGC 602

677 LeuLysMetAspAspGluLysValAlaGluAlaLeuGlnIlePhe 693
|||||
603 CTAAGAATGGAGGATGACAAAGGTAGCAGAGCTGTATTCAAAATTTTAG 652

693 sAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaL 710
|||||
653 AAATACAGGTCACAAATAGAAACAGACCTTCCCAGATACGATCGACCT 702

710 euLeuProValLeuHisHisLysSerLysLysGlyProArgGlnAla 726
|||||
703 TAATTCCTCATTTTACATCAAAAAGCAAGAGGGGTACTCCACACCAAGCA 752

727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGl 743
|||
753 AAACAGGCTGTGCTACTGTATACACGCCCATATTCACAAATAAAGAGTCCA 802

743 nPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnL 760
|||||
803 GCTTGCACAGATTTTGGACCTCAGTAGGAGTCTGAATGCTGATGTCG 852

760 euGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
|||||
853 CAGAACAACTTATACTCCATTTAGTTTTCATGGCCACATTTCTATGTTA 902

777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 793
|||||
903 GCACAGATCAGTTTGTCTCCCAATCTGTAAGTACTAGCAAAATTTTAT 952

793 eValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrL 810
|||||
953 TGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAGAAATGGAA 1002

810 ysLeuTrpValProAspGluGluValSerProGluThrMetValLysIle 826
|||||
1003 AACTGTGGTCTCCAGATGAAGAGGTTTCCCTCTGAAGTACTAGCAAAAGTA 1052

827 GlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnH 843
|||||
1053 CAGGCAATTAACCTTCTGGTAAGGTGGCTTTGGGTATGAAACCAACCA 1102

843 sSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHis 860
|||||
1103 GTCTAAATCTGCCAATTCACCTTCGGTTATTATCAGCGATGTTGGTTA 1152

860 exAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
|||||
1153 GTGAGGTCACCTGCAGAGCAAAAGAGGATCAGTAAATCTGTATGTCT 1202

877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnL 893
|||||
1203 CCCTTCCGATTTAGCTGCTGGTAGTGCCATAATGAAGCTTGTCTCAGGAACC 1252

893 oCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuA 910
|||||
1253 TTGTTTACCATGAATATTATACCCAGAACAGTTTTCAGCTCTGTGCACTTG 1302

910 IaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
|||||
1303 TTATTAATGATCAGTGTACCAAGTAAGGCAGATATTTGCTCAGAAGCTG 1352

927 HisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCy 943

```



```

539 roGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluasp 555
|||||
553 CTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGAT 602
|||||
556 AspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSe 572
|||||
603 GATGAGAAATAAGAAACGAGTTAGAAGTACTTGTAGTCCAAACATGCTC 652
|||||
572 rCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyA 589
|||||
653 CTGCACAGCGTGAAGGTTGTGCGCGTGAATAACTTAAGAAGTTGGCA 702
|||||
589 snProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeu 605
|||||
703 ACCCAACAGCCTACAATCCCTTCTCGAATGATCAAGTTCTCTTG 752
|||||
606 GluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuII 622
|||||
753 GAGAGGATAGCACCCTGTGCATAGATACCGAATCTATCAGTCTCTTAT 802
|||||
622 eLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluaspGluG 639
|||||
803 TAAACAAGTGAACAAATCAATAGATGAACAGCAGATGATGAAGATGAGG 852
|||||
639 lyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysVal 655
|||||
853 GTGTTCCAAGTCAACGATCAGACCATCAGACGAGTCTTGAACCTTAAGTA 902
|||||
656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSe 672
|||||
903 CTCTCATTTACATCCCATCTCATTTTCATTTCTGCTGAAACATTTGAATC 952
|||||
672 rLeuLeuAlaCysLeuLysMetaspGluLysValAlaGluAlaAlaL 689
|||||
953 ATTAAGTGGCTGTCTGAAATGGATGATGAAAGTAGCAGAGCTGCAC 1002
|||||
689 euGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705
|||||
1003 TACAAATTTTCAAAACACAGGAGCAAAATTTGAAGAGATTTTCCACAC 1052
|||||
706 IleArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyPr 722
|||||
1053 ATCAGATCAGCCTGCTCTGCTGTTTACATCACAATCTAAAAAGGACC 1102
|||||
722 oProArg GlnAlaLysTyAlaIleHisCysIleHisAlaIlePheSer 738
|||||
1103 CCCCCTCAGCCCAATATGCCATTCATTTGATCCATGCGATATTTCT 1152
|||||
739 SerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
|||||
1153 AGTAACAGAGACACAGTTTGCACAGATATTTGAGCCTCTGCATAGAGCCT 1202
|||||
755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyH 772
|||||
1203 AGATCCCAAGCAACCTGGAACATCTCATAACACCATTGTTACTATTGGTC 1252
|||||
772 isIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrp 788
|||||
1253 ATATTGCTCTCTTGCACCTGATCAATTTGCTGCTCTTTGAAATCTTG 1302
|||||
789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGl 805
|||||
1303 GTAGCTACTTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCCAGG 1352
|||||
805 yLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluT 822
|||||
1353 GAAAAAGACAACTAACTTTGGGTTCCAGATGAAGAAAGTATCTCTCTGAGA 1402
|||||
822 hrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGly 838
|||||
1403 CAATGGTCAAAATTCAGGCTATTAAATGATGATGTTCCGCTACTTGGC 1452
|||||
839 MetLysAsnAsnHisSerLysSerGlyThrSerThrLeu ArgLeuLeuT 855

```

```

|||||
1453 ATGTAAMTTAMTCACAGTATATCAGGATCTTCTTCCTTAAGSATTTGCTAA 1502
|||||
855 hrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSer 871
|||||
1503 CACAATATATCGCATCGTGTGGGATTTGGMAGAMCAGGGGAAAT...TAG 1549
|||||
872 LysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLy 888
|||||
1550 TAACCAGATATGTCACGCTGAGACTTGTCTGGGAGTCTATTTTGAA 1599
|||||
888 sLeuAlaGlnGluProCysTyHisGluIleIleThrLeuGluGlnTyrg 905
|||||
1600 GCTGGCACAAGAACCCCTGTTACCATGAAATCRTCACAATAGAACAAATATC 1649
|||||
905 InLeuCysAlaLeuAlaIleAsnAspGluCysTyrgGlnValArgGlnVal 921
|||||
1650 AGCTATGCTGCTTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTG 1699
|||||
922 Phe AlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG 938
|||||
1700 TTTGGCCCAAGAACTTCGCAAGGCCCTTCCCGTTTCGGCTTCCACTTG 1749
|||||
938 luTyMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
|||||
1750 AGAATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAGAGGAGAGA 1799
|||||
955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGl 971
|||||
1800 AGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGACGGGA 1849
|||||
971 u 971
:
1850 C 1850

seq_name: gb_pr1:AB014548
seq_documentation_block:
LOCUS AB014548 5177 bp mRNA PRI 06-FEB-1999
DEFINITION Homo sapiens mRNA for KIAA0648 protein, partial cds.
ACCESSION AB014548
VERSION AB014548.1 GI:3327109
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:HJ03994.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5177)
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
DIRECT SUBMISSION
JOURNAL Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnaInfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
2 (sites)
REFERENCE Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
AUTHORS Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
JOURNAL DNA Res 5 (3), 169-176 (1998)
MEDLINE
FEATURES
Location/Qualifiers
1..5177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HJ03994"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"

```



1198 rAspLysArgAspSerAspLeuValArgSerGluLeuGluLysProA 1215  
 |||||  
 3503 TCACAGAGAGACGACTCGATCTTGAAGTCTGAATGGAGAGCCTA 3552  
 1215 rGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMet 1231  
 |||||  
 3553 GAGGAGGAGAAAAACGCCGCTACAGAACAGGAGGAGAAATTAGGTATG 3602  
 1232 AspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnAr 1248  
 |||||  
 3603 GATGACTTGACTAAGTGGTACAGAACAGAAACCTAAAGGAGTCAGCG 3652  
 1248 gSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpP 1265  
 |||||  
 3653 AAGTCGGAAGAGAGGCCATACGCTTCAGAACTCTGATCAACAGCAGTGGC 3702  
 1265 roGluLysArgLysLeuLysGluAspLysLeuGluAsnGluAspGluGln 1281  
 |||||  
 3703 CTGAGGAAAGAGGCTCAAGAGAGATATATTAGAAATGAAGATGAACAG 3752  
 1282 AsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuG1 1298  
 |||||  
 3753 ANTAGTCGCCAAAAGGTTAAAGAGGCCGCCACCAAAACCTCTTGG 3802  
 1298 yGlyThrProLysGluGluProThrMetLysThrSerLysLysGlyS 1315  
 |||||  
 3803 TGCAGGTACACCAAAAGAGAGCCCAATGAATACTCTAAATAAGGAA 3852  
 1315 erLysLysLysSerGlyProProAlaProGluGluGluGluGlu 1331  
 |||||  
 3853 GCATAAAAAAATCTGGACTCCAGCACAGAGGAGGAGGAGAAAGAA 3902  
 1332 ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgVa 1348  
 |||||  
 3903 AGACAAAGTGGAATACGGAACAGAGTCCAAAGACCAACACCGAGT 3952  
 1348 lSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleG 1365  
 |||||  
 3953 GTCAAGGAGAGACAGACAGAGAGAGATCTCCTGAATCTAGTCAATG 4002  
 1365 luSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThr 1381  
 |||||  
 4003 ATCCACACAGTCCACACACAGAGAGGAGGAGGAGGAGGAGGAGGAGG 4052  
 1382 ProSerProSerGlnProLysLysAsnVal 1391  
 |||||  
 4053 CCATCACCATCAACACCAAAAAAATGTC 4082

seq\_name: gb\_pr10:HSU50533

seq\_documentation\_block:  
 LOCUS HSU50533 1852 bp mRNA PRI 27-NOV-1996  
 DEFINITION Human BRCA2 region, mRNA sequence CG008.  
 ACCESSION U50533  
 VERSION U50533.1 GI:1531605  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1852)  
 AUTHORS Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,  
 Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,  
 Farid,L., Frye,C., Hattier,T., Janekci,T., Jiang,P., Kehrer,R.,  
 Leblanc,J.-F., McArthur-Worlison,J., McSweeney,D., Miki,Y.,  
 Peng,Y., Sanson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,  
 Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,  
 Tranchesi,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,  
 Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,  
 Tavtigian,S.V. and Simard,J.  
 Generation of an integrated transcription map of the BRCA2 region  
 on chromosome 13q12-q13

JOURNAL Genomics 36 (1), 86-99 (1996)  
 MEDLINE 96411650  
 REFERENCE 2 (bases 1 to 1852)  
 AUTHORS Simard,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular  
 Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,  
 Quebec City, Quebec G1V 4G2, Canada  
 FEATURES  
 Location/Qualifiers  
 source  
 1. 1852  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /map="13q12-q13"  
 /note="CG008; DSEG numbers: D13S171 and D13S310"  
 BASE COUNT 609 a 348 c 379 g 505 t 11 others  
 ORIGIN

alignment\_scores:  
 Quality: 2953.50 Length: 619  
 Ratio: 4.947 Gaps: 4  
 Percent Similarity: 96.446 Percent Identity: 95.153  
 alignment\_block:  
 US-09-512-581-2 x HSU50533 ..  
 Align seg 1/1 to: HSU50533 from: 1 to: 1852

356 LysValArgSerHisAspProGluAlaIleArgHisAspValIleVa 372  
 |||||  
 3 AAGTGTGAGTTCACATGACCTCGAGGAGCTATTAGACATGATGTTATTGT 52  
 372 lSerIleValThrAlaLysLysAspIleLeuValAsnAspHisL 389  
 |||||  
 53 GTCAATAGTTACAGCTGCTAAAGAGGATATCTCTGTGTCATGATCACT 102  
 389 euLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArg 405  
 |||||  
 103 TACTTTAAATTTTGTGAGAGAGAGAACATTTAGACAAACGATGGAGAGTACGC 152  
 406 LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGl 422  
 |||||  
 153 AAAGAGGCCATGATGGGACTTGCCTCAAAATTTATAAGAAATATGCTTTTACA 202  
 422 nSerAlaAlaGlyLysAspAlaLysGlnIleAlaTyrIleLysAspL 439  
 |||||  
 203 GTCACGACGTGCAAAAGATGCTGCAAAACAGATAGCATGATCAAGACA 252  
 439 ysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArgLeuLeuVal 455  
 |||||  
 253 AATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGT 302  
 456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrCl 472  
 |||||  
 303 GAACGGATCTTTGCTCAATACATGTTCTCACAATTTAGAAACATACAGA 352  
 472 uArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlav 489  
 |||||  
 353 ACGGATGAATGCTTATATTACTTCTATGCCACACTGGGATTTAAATGCTG 402  
 489 alLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGln 505  
 |||||  
 403 TGAAAGCATTTGAATGAATGTGAAATGTCAAAATCTGCTCCGACATCAA 452  
 506 ValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522  
 |||||  
 453 GTPAAGGATTTCTGTCATGATTGATTAAAGCAACCCAAACAGATGCCAGTGT 502  
 522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspP 539  
 |||||  
 503 CAAGGCCATATTTTCAAAAGTGTGTTATTACAAAGAAATTTACCTGATC 552



|||||  
1653 GAAATAACTAAGAAGTTGGCAACCCCAACAGCCTACAAATCCTTTCCT 1702  
598 uGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspT 615  
1703 GGAATGATCAAGTTCTCTGGAGAGGTAGACACCTGGCACATAGATA 1752  
615 hrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGly 631  
1753 CCGAATCTATCAGTGCTCTTAACACAGTGAACAAATCAATAGATGGA 1802  
632 ThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAl 648  
1803 ACAGCAGATGATGAAGATGAGGTGTTCACACUGATCAAGCCATCAGAC 1852  
648 aGlyLeuGluLeuLysValLeuSerPheThrHisProIleSerPheH 665  
1853 AGGCTTGAAGTCTTAAGGTACTCTCATTTACACATCCATCTCATTTTC 1902  
665 isSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAsp 681  
1903 ATTCTGCTGAAACATTTGAATCATTTACTGGCTTGTCTGAAATGGATGAT 1952  
682 GluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLy 698  
1953 GAAAAAGTAGCAGAAGCTGCACFACAAATTTTCAAAAAACACAGGAACAA 2002  
698 sIleGluLysAspPheProHisIleArgSerAlaLeuLeuProValLeuH 715  
2003 AATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGGCTTCCTGTTTAC 2052  
715 isHisLysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHis 731  
2053 ATCACAAATCTAAAAGAGNCCCCCGTCAAGCCAAATATGCCATTCAT 2102  
732 CysIleHisAlaIlePheSerLysLysGluThrGlnPheAlaGlnIlePh 748  
2103 TGTATCCATCGCATATTTCTAGTAAAGAGACCCAGTTTGCACAGATATT 2152  
748 eGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleT 765  
2153 TGAGCCTCTGCATTAAGAGCCTTAGATCCAGCAACCTGGAACATCTCATA 2202  
765 hrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPhe 781  
2203 CACATTTGGTTACTATTGGTCATATTGCTCTCTTGCACCTGATCAATTT 2252  
782 AlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuLe 798  
2253 GCTGCTCCTTTGAATCTTTGGTAGCTACTTTTCATTTGTGAAGATCTTCT 2302  
798 uMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProA 815  
2303 CATGAATGATCGCTTCCAGGAAAAAGACAACTAACTTTGGGTCCAG 2352  
815 spGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMet 831  
2353 ATGAAGAAGTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAATG 2402  
832 MetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyTh 848  
2403 ATGGTTCCGATGGCTACTTGAATGAAAAATAATACAGATAAATCAGAAC 2452  
848 rSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuT 865  
2453 TTCTACCTTAAGATTTGCTACACAAATATTGTCATAGTATGGAGACTTGA 2502  
865 hrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 881  
2503 CAGAACAGGGAAAAATAGTAAACCAAGATATGTCAGCTCTGAGACTTGCT 2552  
882 AlaGlySerAlaIleValLysLeuAlaGlnProCystTyrHisGluIl 898  
|||||

2553 GCTGGAGTGCTATTGTGAAGCTGGCACAGAACCCTGTACCATGAAAT 2602  
898 eIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluC 915  
2603 CATCACATTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGAA 2652  
915 ystYrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931  
2653 GCTATCAAGTAAGACAAAGTGTTCGCCAGAACTTCACAAAGGCCCTTCC 2702  
932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLy 948  
2703 CGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTCCCTTTGTGCAAA 2752  
948 sAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysA 965  
2753 AGATCTGTAAAGGAGAGAGAGCTCATGCTAGGCAATGTTTGGTGAAAA 2802  
965 snIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu 981  
2803 ATATAAATGTAAAGCGGAGTATCTGAAGCAGCATGCAGCTGTAGTGAA 2852  
982 LysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLe 998  
2853 AAATTTATTTGCTCTTACCAGAGTATGTTGTTCCATATACAATTCACCT 2902  
998 uLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuL 1015  
2903 TTTGGCAGCATGACCCAGATTTATGTAAGTACAGGATATTTGAACAACTTA 2952  
1015 ysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLys 1031  
2953 AAGATGTTAAAGAAATGCTTTGTTGTTCTGGAATAATTAATGGCTAAA 3002  
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLy 1048  
3003 AATGAAAAAACAAGTCAAGCTTTTATCAGAAAGATGTTAGAAAAATATA 3052  
1048 sGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysL 1065  
3053 ACAACAAAAGATGCCCAAGGACCAGATGATCAAAAATGAATGAAAAAC 3102  
1065 euTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThr 1081  
3103 TGTACACTGTGTGTGATGTGGCATGAATATCATCATGTCAAGAGTACT 3152  
1082 ThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePh 1098  
3153 ACATACAGTTTGAATCTCTCTAAAGACCCGGTACTACCAGCTCGTTTCTT 3202  
1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProG 1115  
3203 CACTCAACCTGCACAAAGATTTTCAGTAACACCAAAAATTTATCTGCCTCTG 3252  
1115 luMetLysSerPheThrProGlyLysProLysProLysThrThrAsnValLeu 1131  
3253 AAATGAATCATTTTTTCATCTCTGGAACCACTTAAACACCAATGTTCTA 3302  
1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148  
3303 GGAGCTGTTTAAACAAGCACCTTTTCATCAGCAGGCAAGCAATCTCAGACCA 3352  
1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProS 1165  
3353 ATCATCAGCAATCGAAACTGTAAAGCAATGCAAGCAGCAGCTCAATCCAA 3402  
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHis 1181  
3403 GCTCTCTGGAAAGAAATAAGGGAGGCTTGTAGTCTGAAATGGATCAC 3452  
1182 SerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSe 1198  
3453 AGTGAATAATGAAGATTACACAATGTCTTCACCTTTGCGGGGAAAAAAG 3502



## ORIGIN

alignment\_scores:  
Quality: 6994.00 Length: 1360  
Ratio: 5.154 Gaps: 0  
Percent Similarity: 99.779 Percent Identity: 99.706

alignment\_block:  
US-09-512-581-2 x AB023196 ..

Align seg 1/1 to: AB023196 from: 1 to: 5110

32 ValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAs 48  
|||||  
3 CTTGAGACGATTAAAGATGCTTGTGAAACCTTTTATGGATATGACACGGA 52  
48 pSerGluGluGluLysGluLeuLysLeuLysLeuLysLeuLysLeuLys 65  
|||||  
53 CTTGGAAGAAAGAGAGCTTTATTTAAACCTAGCTTTTACATCTCTGCTT 102  
65 exAspPheLeuLysHisProGlyLysAspValArgLeuLeuValAla 81  
|||||  
103 CAGATTTTTTCTCAGCATCTCTGATAAGATGTTTCGCTTACTGGTAGCC 152  
82 CysCysLeuAlaAspPheArgIleTyrAlaProGluAlaProTyrTh 98  
|||||  
153 TGTCTGCTTGTGATATTTTACAGATTTATGCTCTGAAAGCTCTTACAC 202  
98 rSerProAspLysLeuLysAspPhePheMetPheIleThrArgGlnLeu 115  
|||||  
203 ATCCCTGATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGA 252  
115 ysglyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeu 131  
|||||  
253 AGGGCTAGAGGATACAAAGGCCACAGTTCAATAGGTATTTTATTTA 302  
132 LeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuG 148  
|||||  
303 CTTGAGAACATTGCTGGGTCAAGTCATATAACATATGCTTTGAGTTAGA 352  
148 uaspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerVal 165  
|||||  
353 AGATAGCAATGAAATTTTCCACCAGCTATACAGAACCTTTATTTTCAGTTA 402  
165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181  
|||||  
403 TAAACATGGCCACATCAGAAGTCCATATGCACATGGTAGACCTTATG 452  
182 SerSerIleileCysGluGlyAspThrValSerGlnGluLeuLeuAspTh 198  
|||||  
453 AGCTCTATTATTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATAC 502  
198 rValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAla 215  
|||||  
503 GGTTTAGTAAATCTGGTACCTGCTCATAGAATTTTAAACAAGCAAGCAT 552  
215 yAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluPro 231  
|||||  
553 ATGATTTGGCAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCA 602  
232 TyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerI 248  
|||||  
603 TATATTACCAATTTTTTAAATCAGGTTCTGATGCTTGGGAAACATCTAT 652  
248 eSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsn 265  
|||||  
653 CAGCGATTTGTCAGACATGCTCTTGAACCTTAATTTTGGAGCTCTACAATA 702  
265 leAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeu 281  
|||||  
703 TTGATAGTCTATGCTGCTCTCTGTTTTACCACGAGCTTGAATTTAAATTA 752

282 LysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAla 298  
|||||  
753 AAGAGCAATGATAATGAGCAGCGCTTACAAGTTGTTAAACTACTGCGAAA 802  
298 smetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeu 315  
|||||  
803 AATGTTTGGGCAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTT 852  
315 rpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeu 331  
|||||  
853 GGCAGTGCTACTTGGCAGGTTTAAATGATATCCATGTACCAATCCGCTG 902  
332 GluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeu 348  
|||||  
903 GAATGTGTCAAAATTTGCTAGCCATTTGCTCATGAACCATCTGATTTAG 952  
348 alyAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGlu 365  
|||||  
953 AAAAGACTTAAACAGAGTATCTTAAAGTGAGGTACATGACCTTGAGGAG 1002  
365 laIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAsp 381  
|||||  
1003 CTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGCTAAAAGGAT 1052  
382 IleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLe 398  
|||||  
1053 ATTCTTCTGCTCAATGATCATTACTTAATTTTGTGAGAGAGAGAACATT 1102  
398 uaspLysArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGln 415  
|||||  
1103 AGACAACCATGGAGAGTACGCAAGAGCCATGATGGGACTTGGCCAAA 1152  
415 letyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLys 431  
|||||  
1153 TTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAGAGATGCTGCAAAA 1202  
432 GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSe 448  
|||||  
1203 CAGATAGCATGGATCAAGACAAATTTGCTACATATATATATCAAAATAG 1252  
448 rIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVal 465  
|||||  
1253 TATGTATGATCGACTACTTGTGAACGGATCTTTGCTCAATACATGGTTC 1302  
465 roHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyr 481  
|||||  
1303 CTCACAATTTAGAACTACAGAAGCGATGAATGCTTATATTACTTGTAT 1352  
482 AlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCy 498  
|||||  
1353 GCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAATG 1402  
498 sGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuLeuIleLysG 515  
|||||  
1403 TCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGC 1452  
515 lnProLysThrAspAlaSerValLysAlaIlePheSerLysValMetVal 531  
|||||  
1453 ACCCAAAACAGATGCCAGTCAGGCCATATTTTCAAAAGTGATGGTT 1502  
532 IleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLy 548  
|||||  
1503 ATTACAAGAAATTTTACCTGATCTCTGTAAGGCTCAGGATTTTCATGAAGA 1552  
548 sPheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluV 565  
|||||  
1553 ATTCACACAGGTTTAGAAGATGATGAGAAAATAAGAAAGCAGCTTAGAAG 1602  
565 alLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArg 581  
|||||  
1603 TACTTGTAGTCCACATGCTCTCTGCAACGAGCTGAAAGTTGTGTGCGT 1652  
582 GluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLe 598



```

3316 GTTTGGAACTCTCTAAAGACCGGTACTACAGCTCGTTTCTTCACRCAA 3365
1101 ProAspLysAsnPhSerAsnThrLysAsnThrLeuProGluMetLys 1117
|||||
3366 CCTGACAAAGATTCAGTAACACCAAAATATCTGCTCTCTGGAATGAA 3415
1117 sSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlav 1134
|||||
3416 ATCATTTTTCACCTCTGGAAACCTAAACCAACCAATGTTCTAGGAGCTG 3465
1134 alAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer 1150
|||||
3466 TTAAACAGCCACTTTCATCAGCAGCAGCAATCTCAGACCAAAATCATCA 3515
1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerPr 1167
|||||
3516 CGAATGGAACCTGTAGGCAATGCAAGCAGCAGCTCAATCAACAGCTCTCC 3565
1167 oGlyArgTleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA 1184
|||||
3566 TGGAGAAATAAGGGAGGCTTGATAGTTCTGCAATGATGATCAGTCAAG 3615
1184 snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
|||||
3616 ATGAAGATTACACAATGCTTTCACCTTTGCCGGGAAAGAGTACAAG 3665
1201 ArgAspSerSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
|||||
3666 AGAGCAGCAGCTGTATCTGTGAAGCTCTCAATTTGGAGAGGCTAGAGCAG 3715
1217 gLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspL 1234
|||||
3716 AAAAAAACCCCTCAGAACCGGAGGAGAAATTAGGTATGATGACT 3765
1234 euThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg 1250
|||||
3766 TGACTAAGTTGGTACAGAACAGACAGACTAAAGGCAGTCAGCGAAGTCGG 3815
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluG 1267
|||||
3816 AAAAGAGCCATACGGCTTCAGAAATCTGATGACAGCAGTGCGCTGAGGA 3865
1267 uLysArgLeuLysGluAspLleLeuGluAsnGluAspGluGlnAsnSerp 1284
|||||
3866 AAAGAGCTCAAGAACAGATATATTAGAAATGAAGATGAACAGATAGTC 3915
1284 roProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
|||||
3916 CGCCAAAAAGGGTAAAAAGAGCGCGACCAACCAACCTCTTGGTGGAGGT 3965
1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317
|||||
3966 ACACCAAAAGAGGAGCCCAACCAATGAAACCTCTAAAAAGAGCAACAAA 4015
1317 sLysSerGlyProProAlaProGluGluGluGluGluGluArgGlns 1334
|||||
4016 AAATCTGGACCTCCAGCACAGGAGGAGGAGAGAGAGAGAGAGAGACAAA 4065
1334 erGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1350
|||||
4066 GTGGAATATACGAACAGACAGAAAGTCCAAAAAGCAACAGCAGCGAGTGTCAAGG 4115
1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367
|||||
4116 AGAGCAGCAGAGAGAGAGAAATCTTCAATCTAGTCAATTTGAATCCAC 4165
1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerp 1384
|||||
4166 ACAGTCCACACACAGAGAGGAGGAGGAGAGACCATCAAAAAAGCCATCAC 4215
1384 roSerGlnProLysLysAsnVal 1391
|||||
4216 CATCACACCAAAAAAAATGTG 4238

```

seq\_name: gb\_pri:AB023196

seq\_documentation\_block:

LOCUS AB023196 5110 bp mRNA PRI 16-JUN-1999  
 DEFINITION Homo sapiens mRNA for KIAA0979 protein, partial cds.  
 ACCESSION AB023196  
 VERSION AB023196.1 GI:4589601

KEYWORDS

SOURCE

Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII  
 SK plus clone:hj07056.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE



484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500  
1516 TGGATTTAAATGCTGTAAGCAATGCAATGAAATGTGGAATGTCAAAAT 1565  
501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeuLeuLeuLeu 517  
1566 GTGCTCCGACATCAAGTAAGGATTTGCTTGACCTGATTAAAGCAACCCAA 1615  
517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534  
1616 AACAGATGCCAGTGTCAAGCCATATTTTCAAAAGTGATGGTTATTACAA 1665  
534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550  
1666 GAAGTTTACCTGATCTGTGAAGGCTCAGGATTTTCATGAAGAAATTCACA 1715  
551 GlnValLeuGluAspGluLysIleArgLysGlnLeuGluValLeuVa 567  
1716 CAGGTGTTAAGAGATGATGAGAAATTAAGAAAGCAGTTAGAAGTACTTGT 1765  
567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584  
1766 TAGTCCAACATGCTCTCAAGCAGGCTGAAGGTTGTGTGCGFGAAATAA 1815  
584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600  
1816 CTAAGAAGTTGGCAACCCCAACAGCCCTACAAATCTTTCTCTGGAATG 1865  
601 lIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrCluSe 617  
1866 ATCAAGTTTCTCTTGAGAGGATAGCAGCTGTGCACATAGATACCGAATC 1915  
617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634  
1916 TATCAGTGTCTTATTAAACAAGTGAACAATCAATAGATGAACAGCAG 1965  
634 sPaspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650  
1966 ATGATGAAGATGAGGTGTCTCAACTGTATCAAGCCATCAGACAGGTCTT 2015  
651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667  
2016 GAATGCTTTAAGGTACTCTATTACACATCCCATCTCATTTCAATCTCTC 2065  
667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684  
2066 TGAACATTTGAATCATTTACTTGCCCTGTCTGAAATGGATGATGAATAAG 2115  
684 aAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700  
2116 TAGCAGAAGCTGCATACAAAATTTTCAAAAACACAGGAAGCAAAATTTAA 2165  
701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIle 717  
2166 GAGGATTTTCCACATCAGATCAGCTTGTCTCTGTTTTATACATCACAA 2215  
717 sSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleH 734  
2216 ATCTAAAAAAGGACCCCGCTCAGCCAAATATGCCAATATCATTTGTATCC 2265  
734 isAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750  
2266 ATCCGATATTTTCTAGTAAGAGGCCAGCTTTGACAGATATTTGAGCCT 2315  
751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767  
2316 CTGCATAAGAGCCTAGATCCAGCAACCTTGGACATCTCATACACCAAT 2365  
767 uValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaP 784  
2366 GGTTACTATTGGTCATATTTGCTCTCTTGCACCTGATCAATTTGCTGCTC 2415  
784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800

2416 CTTTGAAATCTTTGGTAGCTACTTTTCAATGTGAAAGATCTTCTCATGAAT 2465  
801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluG 817  
2466 GATCGCTTCCAGGAAAGAAACAACAACTTTGGGTTCCAGATGAAGA 2515  
817 uValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834  
2516 AGTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGTTTC 2565  
834 rGTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850  
2566 GATGCTACTTGGAAATGAAAAATAATACAGTAAATCAGGAACCTTCTACC 2615  
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluG 867  
2616 TTAAATGCTTAACAACATATTTGCATGATGGAGACTTCACAGAACA 2665  
867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884  
2666 GGGCAAAATTTAGTAAACACAGATATGTACGCTCTGAGACTTGTCTGGGA 2715  
884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900  
2716 GTGCTATTGTGAAGCTGGCACAAGAACCCCTGTACCATGAAATCATCA 2765  
901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCystyrG 917  
2766 TTAAACAACATATCAGCTATGCTGATGCTATCAACGATGATGCTATCA 2815  
917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934  
2816 ACTAAGACAAGTGTGTGCCAGAAACTTCACAAAGGCCCTTCCCGTTTAC 2865  
934 rGLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950  
2866 GGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCT 2915  
951 ValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967  
2916 CTAAAGGAGAGAAGACCTCATCTAGGCAATGTTGGTCAAAAATATAA 2965  
967 nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeu 984  
2966 TGTAAAGCGGGAGTATCTGAAGCAGCATGCTGTAGTGTAGTAAAAATAT 3015  
984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000  
3016 TGTCTCTTCTACAGAGTATGTTTCCATATACAAATTCACCTTTTGGCA 3065  
1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa 1017  
3066 CATGACCCAGATTTATGCAAGTACAGGATATTGAACAACCTTAAAGATCT 3115  
1017 lLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluA 1034  
3116 TAAACAATGTCTTGGTTTGTCTTGGAAATATTAATGGCTGCTAAAAATGAA 3165  
1034 snAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr 1050  
3166 ATACAGTCCAGCTTTTATCAGAAAGATGTAAGAAATATTAAAGCAACA 3215  
1051 LysAspAlaGlnGlyProAspAlaLysMetAsnGlnLysLeuIleThr 1067  
3216 AAAGATGCCCAAGGACCATGATGCAAAATCAATGAAAACTGTACAC 3265  
1067 rValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrS 1084  
3266 TGTGTGTGATGTTGTCATGAATATCATCATGTCAAGAGTACTATACATA 3315  
1084 erLeuLeuSerProLysAspProValLeuProAlaArgPhePheThrGln 1100



```

/number=29
3438..3583
/gene="AS3"
/number=30
3584..3689
/gene="AS3"
/number=31
3690..4129
/gene="AS3"
/number=32
4130..4354
/gene="AS3"
/number=33
4355..4810
/gene="AS3"
/number=34
1755 a 944 c 1074 g 1480 t

BASE COUNT
ORIGIN

alignment_scores:
    Quality: 7101.00      Length: 1391
    Ratio: 5.131          Gaps: 0
    Percent Similarity: 99.497    Percent Identity: 98.994

alignment_block:
US-09-512-581-2 x HSU95825 ..

Align seg 1/1 to: HSU95825 from: 1 to: 5253

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17
|||||
66 ATGGCTCATTCAGAGCTAGGACCAATGAGAAATACATATCGCC 115
|||||
17 oGlyValLysGluLeuSerAspLysIleSerLysGluMetValArgA 34
|||||
116 TGGGGTCAAGGAATATACAGATAAAATATCTAAAGAGAGATGGTGAGAC 165
|||||
34 rGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
|||||
166 GATTAAAGATGGTGTGAAACCTTTTATGGATATGGACAGGACTCTGAA 215
|||||
51 GluGlyLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
|||||
216 GAAGAAAAGAGAGCTTTATTTAAACCTAGCTTTACATCTTGCCTCAGATT 265
|||||
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84
|||||
266 TTTTCTCAAGCATCCTGATAAAGATGTCGCTTACTGGTAGCCTGCTGCC 315
|||||
84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
|||||
316 TTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCT 365
|||||
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
|||||
366 GATAAACTAAGAGATATATTTATGTTTATAACAGACAGTTGAAGGGCT 415
|||||
117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134
|||||
416 AGAGGATACAAAGAGCCACAGTTCAATAGTATTTTATTTGCTTGAGA 465
|||||
134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150
|||||
466 ACATTGCTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGC 515
|||||
151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167
|||||
516 AATGAAATTTTACCAGCTATACAGAACCTTATTTTCACTTATAACAA 565
|||||
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184
|||||
566 TGGCCACAATCAGAAAGTCCATATGCATGTCATGAGACCTTATGAGCTCTA 615

```

```

184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
|||||
616 TTATTTGTGAAGGTGATACAGTGTCTCAGAGCTTTTGGATACGGTTTA 665
|||||
201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217
|||||
666 GTAAATCTGGTCTGCTCATAGAATTTAAACAAGCAAGCATATGATT 715
|||||
217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234
|||||
716 GGCAAGAGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTA 765
|||||
234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250
|||||
766 CCAATTTTAAATCAGCTTCTGATGCTTGGGAAAACATCTATCAGCGAT 815
|||||
251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267
|||||
816 TTGTCAGAGCATGTCTCTGACTTAATTTTGGAGCTCTACAATATTGATAG 865
|||||
267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284
|||||
866 TCATTTGCTGCTCTGCTTTTACCCAGCTTGAATTTAAATTAAGAGCA 915
|||||
284 snAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
|||||
916 ATGATAATGAGAGCGGCTCAAGTTGTTAAACTACTGGGAAAATGTTT 965
|||||
301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCy 317
|||||
966 GGGCAAGAGATTCAAGATTTGGCTTCTCAAAACAAGCCACTTTTGGCAGTG 1015
|||||
317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334
|||||
1016 CTACTTGGGAGGTTTAAATGATATCCATGTACCATCCGCTTGGAAATGTG 1065
|||||
334 allysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
|||||
1066 TGAATTTGCTAGCCATTGCTCATGAACCATCTGATTTAGCAAAAGAC 1115
|||||
351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
|||||
1116 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTCGAGGAAGCTATTAG 1165
|||||
367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
|||||
1166 ACATGATGTTATTGTGTCAATAGTTACAGCTGTCTAAAAGGATATTCTTC 1215
|||||
384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
|||||
1216 TGGTCAATGATCACTTACTTAATTTTGGGAGAGAGAACATTAGACAAA 1265
|||||
401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
|||||
1266 CGATGGAGAGTACGCAAGAGCCATGATGGACTTGCCTCAAAATTTATAA 1315
|||||
417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434
|||||
1316 GAAATATGCTTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAG 1365
|||||
434 laTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450
|||||
1366 CATGGATCAAGAGACAAATTTGCTACATATATATATCAAAATAGATTGAT 1415
|||||
451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467
|||||
1416 GATCGACTACTTGTGTAACGGATCTTTTGCCTCAATACATGTTGTTCTCACA 1465
|||||
467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484
|||||
1466 TTTAGAAAATACAGAACGGATGAATGCTTATATATTACTTGTATGCCACAC 1515

```







```
884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThr 900
|||||
2777 GTGCTATTGTGAAGCTGGCAAGAACCCCTGTACCATGAATCATACA 2826
|||||
901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG1 917
|||||
2827 TTAGAACAATATCACTATGCTATGCTATGCTATCAACGATGAATGCTATCA 2876
|||||
917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
|||||
2877 AGTAAGACAAGTGTTCGCCAGAACCTTCACAAAGGCCCTTTCCCGTTTAC 2926
|||||
934 rgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950
|||||
2927 GGCCTCCACITGACTATATGCAATCTGTGCCCTTTGTGCAAAAGATCCT 2976
|||||
951 ValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
|||||
2977 GTAAAGGAGAGAGAGAGCTCATGCTTAGGCAATGTTTGGTGAAAAATATAAA 3026
|||||
967 nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL 984
|||||
3027 TGTAAAGCGGGAGTATCTGAAGCAGCATGCTGTAGTGAATAATAT 3076
|||||
984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
|||||
3077 TGTCTCTTACACAGATGATGTTGTTCCATATACAAATTCACCTTTTGCA 3126
|||||
1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa 1017
|||||
3127 CATGACCCAGATATTGTCAAAGTACAGGATATTCAACAACCTTAAAGATGT 3176
|||||
1017 LysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGluA 1034
|||||
3177 TAAAGAATGCTTTGGTGTGTTCTGGAATATTAAATGGCTAAAAATGAAA 3226
|||||
1034 snAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr 1050
|||||
3227 ATAACAGTCAGCCTTTATCAGAAGATGGTAGAAATATATAACAACA 3276
|||||
1051 LysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrTh 1067
|||||
3277 AAAGATGCCAAGGACCAAGATGATGCAAAAATGAATGAAAACTGTACAC 3326
|||||
1067 rValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrS 1084
|||||
3327 TGTGTGTGATGTTGCCATGAATATCATCATGTCAAGAGTACTACATACA 3376
|||||
1084 erLeuGluSerProLysAspProValLeuProAlaAaGpPheThrGln 1100
|||||
3377 GTTTGGAATCTCCTTAAGACCCGGTACTACCAGCTCGTTCTTCACATCA 3426
|||||
1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLy 1117
|||||
3427 CCTCACAGAATTTTCAGTAACACCAAAAATATCTGCTCTCGTGAATGAA 3476
|||||
1117 sSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlav 1134
|||||
3477 ATCATTTTCTACTCCTGGAAACCTAAACCAACCAATGTTCTTAGGAGCTG 3526
|||||
1134 alAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer 1150
|||||
3527 TTAACAGCCACTTTTCATCAGCAGCAAGCAATCTCAGACCAATCATCA 3576
|||||
1151 ArgMetGluThrValSerAsnAlaSerSerSerSerSerAsnProSerSerPr 1167
|||||
3577 CGAATGGAACCTGTAAAGCAATGCAAGCAGCAGCTCAATCAACAGCTCTCC 3626
|||||
1167 oGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA 1184
|||||
3627 TGGAGAAATAAAGGGAGGCTTGATGTTCTGAATGATGATCAGCTGAAA 3676
|||||
1184 snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
|||||
```

```
|||||
3677 ATGAAGATTACACAATGCTTTCACCTTTGCCGGGAAAAAAGTGACAAG 3726
|||||
1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
|||||
3727 AGAGACGACTCTGATCTTGTAAAGTCTGAATTTGGAGAAGCCCTAGAGCAG 3776
|||||
1217 qLysLysThrProValThrGluGlnGluLysLeuGlyMetAspAspL 1234
|||||
3777 GAAAAAACGCCCTGCAGAACAGGAGGAGAAATTAGGTATGGATGACT 3826
|||||
1234 euThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg 1250
|||||
3827 TGACTAAGTTGGTACAGAACAGAAACCTTAAAGGCAGTCACGGAATCGG 3876
|||||
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluG1 1267
|||||
3877 AAAAGAGGCCATACGGCTTCAGAAATCTGATGAACAGCAGTGGCCTGAGGA 3926
|||||
1267 uLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerP 1284
|||||
3927 AAAGAGCTCAAGAAGATATATTAGAAATGAAGATGAACAGAAATAGTC 3976
|||||
1284 roProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
|||||
3977 GCGCAAAAAGGGTAAAGAGCGCCGACCACCAAAACCTCTTGGTGGAGGT 4026
|||||
1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317
|||||
4027 ACACCAAAAGAGAGCCAAACAAATGAAACCTTCTAAAAAGGAAGCAAAA 4076
|||||
1317 sLysSerGlyProProAlaProGluGluGluGluGluGluGlnS 1334
|||||
4077 AAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAAGACAAA 4126
|||||
1334 erGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1350
|||||
4127 GTGCAATACGGAACAGAAAGTCCAAAAGCAACAGCAGCAGGTGTCAAGG 4176
|||||
1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367
|||||
4177 AGACACAGCAGAGAGCAGATCTCTCTGAATCTAGTGAATTTGAATCCAC 4226
|||||
1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP 1384
|||||
4227 ACAGTCACACCCACAGAAAGCAGGAGGAGGAGGACCATCAAAAACGCCATCAC 4276
|||||
1384 roSerGlnProLysLysAsnVal 1391
|||||
4277 CATCACAAACCAAAAAAATGTC 4299
|||||
```

seq\_name: gb\_pr10:HSU95825

```
seq_documentation_block:
LOCUS      HSU95825      5253 bp      mRNA      PRI      05-APR-1999
DEFINITION Human androgen-induced prostate proliferative shutoff associated
              protein (AS3) mRNA, complete cds.
ACCESSION  U95825
VERSION    U95825.2  GI:4559409
KEYWORDS   human.
```

```
ORGANISM    human.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 5253)
AUTHORS     Geck,P., Szelel,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
TITLE       Androgen-induced proliferative shutoff in prostate cancer cells
JOURNAL     Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
REFERENCE   2 (bases 1 to 5253)
AUTHORS     Geck,P., Szelel,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
TITLE       Early gene expression during androgen-induced inhibition of
             proliferation of prostate cancer cells: a new suppressor candidate
             on chromosome 13, in the BRCA2-Rb1 locus
```



927 TCATTGCTGCTCTCTGTTTTACCCAGGCTTGAAATTAATTAAGAGCA 976  
 284 snAspAsnGluGluArgLeuGlnValValLysLeuAlaLysMetPhe 300  
 977 ATGATAATCAGGAGCGCTTACAAGTTGTTAACTACTGCCAAAATGTTT 1026  
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317  
 1027 GGGGCAAGGATTCAAGATTGCTTCTCAAAACAAGCCACTTTGGCAGTG 1076  
 317 sTrpLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334  
 1077 CTACTTGGCAGGTTTAATGATATCCATGATACCAATCCGCCCTGGAAATG 1126  
 334 aLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350  
 1127 TGAATTTGCTAGCCATTGCTCATGAACCATCTTGATTTAGCAAAAGAC 1176  
 351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367  
 1177 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTCGAGGAAGCTATTAG 1226  
 367 qHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384  
 1227 ACATGATGTTATTTGTCAATAGTTACAGCTGCTAAAAAGGATATCTTTC 1276  
 384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400  
 1277 TGGTCAATGATCATTACTTAATTTTGTGAGAGAGAACATTAGACAAA 1326  
 401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417  
 1327 CGATGAGAGTAGCGCAAAAGAACCCATGATGGGACTTGGCCAAATTTATAA 1376  
 417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434  
 1377 GAAATATGCTTTACAGTAGCAGCTGGAAGAAGATGCTGCAAAACAGATAG 1426  
 434 IaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450  
 1427 CATGGATCAAGACAAAATGCTACATATATATATATCAAAAATAGTATTGAT 1476  
 451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467  
 1477 GATCGACTACTTGTGTAAGCGGATCTTGTCTCAATACATGGTTCTCCACAA 1526  
 467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484  
 1527 TTTAGAACTACAGAACGGATGAATGCTTATATTTACTTGTATGCCACAC 1576  
 484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500  
 1577 TGGATTTAAATGCTGTGAAGCAATGAATGAATGTGAATGTGAAATGTCAA 1626  
 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProLy 517  
 1627 CTGCTCCGACATCAAGTAAGGATTTGCTTGACTTGATTAAGCAACCCAA 1676  
 517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534  
 1677 AACAGATGCCAGTCAAGGCCATATTTCAAAAGTGATGGTTATTACAA 1726  
 534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550  
 1727 GAAATTTACCTGATCTGTTAAGGCTCAGGATTTATGAAGAAATTCACA 1776  
 551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567  
 1777 CAGGTGTTAGAAGATGATGAGAAAATAAAGAAAGCAGTTTAGAAGTACTTGT 1826  
 567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584  
 1827 TAGTCCAACATGCTCTCTGCAAGCAGGCTGAAGGTTGTGTCGCTGAATAA 1876

584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600  
 1877 CTAACAAGTTGGCAACCCCAACAGCCTTACAAATCCTTTCTCTGGAATG 1926  
 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSe 617  
 1927 ATCAAGTTTCTCTGGAGAGGATAGCCTGTGCACATAGATACCGAATC 1976  
 617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634  
 1977 TATCAGTCTCTTATTAAACAAGTGAACAAATCAATAGATGAACAGCAG 2026  
 634 sPAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650  
 2027 ATGATGAAGATCAGGCTGTCCAACATGATCAAGCCATCAGACAGGCTCT 2076  
 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667  
 2077 GAACGTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCAFTTGC 2126  
 667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684  
 2127 TGAACATTTGAATCATTTACTGCTTGTCTGAAATGGATGATGAAAAAG 2176  
 684 aIaAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700  
 2177 TAGCAGAAGCTGCATCAAAATTTTCAAAACACAGGAAGCAAAATTTGAA 2226  
 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLy 717  
 2227 GAGGATTTTCCACATCATCAGATCAGCCTTGTCTGTTTATACATCACAA 2276  
 717 sSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleH 734  
 2277 ATCTAAAAAGGAGCCCGCTCAAGCCAAATATGCCATTCATTGTATCC 2326  
 734 IsAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750  
 2327 ATGCGATATTTTCTAGTAAAGAGAGCCAGTTTGCACAGATATTTGAGCCT 2376  
 751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767  
 2377 CTGCATAGAGCCTAGATCCAAGCAACCTTGGAAACATCTCATACACCAAT 2426  
 767 vUalThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlap 784  
 2427 GCTTACTATTGCTCATATTGCTCTCTTGCACCTGATCAATTTGCTGCTC 2476  
 784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800  
 2477 CTTTGAATCTTTTGGTAGCTACTTTCATTTGTGAAAGATCTTCTCATGAAT 2526  
 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGl 817  
 2527 GATCGCTTCCAGGGAAGAAAGCACTAAACTTTGGTTCCAGATGAAGA 2576  
 817 vUalSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834  
 2577 ACTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGGTTC 2626  
 834 rGTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850  
 2627 GATGCTACTTGGAAATGAATAATAATACAGTAAATCAGGAACCTTCTTACC 2676  
 851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGl 867  
 2677 TTAAGATGCTAACCAACAATATTGCATAGTGGAGACTTGCACAGAACA 2726  
 867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884  
 2727 GGGGAAATTTAGTAAACCAGATATGTACGCTGAGACTTGTGCTGGGA 2776



exon 3645..3750  
/number=31  
exon 3751..4190  
/number=32  
misc\_feature complement(3864..4286)  
misc\_feature /note="matches EST A1680124 from clone IMAGE:2264521"  
4001..4119  
misc\_feature /note="matches EST N83765 from clone KK2745"  
4135..4450  
exon /note="matches EST F12354 from clone c-38h11."  
4191..4298  
/number=33  
misc\_feature 4255..4307  
exon /note="matches EST T64867 from clone 22264"  
4299..4434  
/number=34  
misc\_feature join(4398..4533,4518..4660,4659..4675,4674..4693)  
exon /note="matches EST H44781 from clone 188032"  
4435..7444  
/number=35  
misc\_feature 4537..5022  
misc\_feature /note="matches EST A1681422 from clone IMAGE:2272632"  
4537..4892  
misc\_feature /note="matches EST AA911125 from clone IMAGE:1416862"  
4538..4969  
misc\_feature /note="matches EST A1275866 from clone IMAGE:1877482"  
4538..4947  
misc\_feature /note="matches EST AA987361 from clone IMAGE:1603350"  
4538..4943  
misc\_feature /note="matches EST A1299646 from clone IMAGE:1898010"  
4562..5003  
misc\_feature /note="matches EST AA447433 from clone 784555"  
4571..4854  
misc\_feature /note="matches EST H93424 from clone 220700"  
join(4573..4649,4650..4815)  
misc\_feature /note="matches EST W20070 from clone 306049"  
complement(4629..4937)  
misc\_feature /note="matches EST AW089775 from clone IMAGE:2593146"  
4746..4994  
misc\_feature /note="matches EST AA747568 from clone IMAGE:1269696"  
complement(4803..5332)  
misc\_feature /note="matches EST A1911784 from clone IMAGE:2329873"  
complement(4828..5296)  
misc\_feature /note="matches EST AA563884 from clone IMAGE:1013990"  
4861..5243  
misc\_feature /note="matches EST D20082 from clone pml283."  
complement(join(4872..4921,5281..5332))  
misc\_feature /note="matches EST AW212954 from clone IMAGE:2647801"  
join(4872..4921,5281..5371)  
misc\_feature /note="matches EST A1177347 from clone 621810"  
complement(4872..4921)  
misc\_feature /note="matches EST AW048929 from clone  
UI-M-BH1-amo-c-09-0-UI"  
4872..4921  
misc\_feature /note="matches EST A1154932 from clone 599194"  
complement(4876..5333)  
misc\_feature /note="matches EST A1016896 from clone IMAGE:1627886"  
complement(join(4922..4975,5281..5332))  
misc\_feature /note="matches EST A1012649 from clone RPLBA02"  
complement(4922..4975)  
misc\_feature /note="matches EST AA899586 from clone  
UI-R-E0-cy-e-02-0-UI; matches EST A1172172 from clone  
RMUBV82"  
complement(4931..5332)  
misc\_feature /note="matches EST AA625961 from clone 745478"  
complement(4951..5332)  
misc\_feature /note="matches EST A1140790 from clone IMAGE:1566222"  
complement(5021..5309)  
misc\_feature /note="matches EST H45693 from clone 188292"

alignment\_scores:  
Quality: 7155.00 Length: 1391  
Ratio: 5.155 Gaps: 0

Percent Similarity: 99.784 Percent Identity: 99.712  
alignment\_block:  
US-09-512-581-2 x HS26H231 ..  
Align seg 1/1 to: HS26H231 from: 1 to: 7444  
1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyProPr 17  
127 ATGGCTCATTCAAAGACTAGACCAATGATGGAAAAATTACATATCGGCC 176  
17 OGlyValLysGluIleSerAspLysIleSerLysGluMetValArgA 34  
177 TGGGGTCAAGGAATAATACAGATAAAATATCTAAAGAGAGATGTTGAGAC 226  
34 rGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50  
227 GATTAAGATGGTGTGAAAACCTTTTATGGATATGGACGAGACTCTGAA 276  
51 GluGluLysGluLeuTyLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67  
277 GAAGAAAGAGAGCTTATTTAAACCTAGCTTTACATCTTGCTTCAGATT 326  
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84  
327 TTTTCTCAAGCATCTCTGATAAAGATGTTCTGCTTACTGTAGCTGTGCC 376  
84 euAlaAspIlePheArgIleTyAlaProGluAlaProTyThrSerPro 100  
377 TTGCTGATATTTTTCAGGATTTATGCTCTGAAAGCTCTTACACATCCCT 426  
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117  
427 GATAAACTAAAGAGATATATTATGTTTAAACAAGACAGTTTGAAGGGCT 476  
117 uGluAspThrLysSerProGlnPheAsnArgTyPheTyIleuLeuGluA 134  
477 AGAGGATACAAAGAGCCCAATTCATAGTATTTTATTTACTTGAGA 526  
134 snIleAlaTrpValLysSerTyTrasnIleCysPheGluLeuGluAspSer 150  
527 ACATTGCTGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGATAGC 576  
151 AsnGluIlePheThrGlnLeuTyArgThrLeuPheSerValIleAsnAs 167  
577 AATGAAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATATAACA 626  
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184  
627 TGGCCCAATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTA 676  
184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200  
677 TTTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGTTTA 726  
201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyAspLe 217  
727 GTAATCTGGTACCTGCTCATAGAATTTAAACAAGCAAGCATATGATTT 776  
217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyIleT 234  
777 GGCAAGGCTTTACTGAAGAGGAGAGCTCAAGCTATTGAGCCATATATTA 826  
234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250  
827 CCAATTTTAAATCAGGTTCTGATGCTTGGGAAAAACATCTATCAGCGAT 876  
251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyTrasnIleAspSe 267  
877 TTGTACAGAGCATGCTTTGACTTAATTTTGGAGCTCTACATATTTGATAG 926  
267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284



gb\_hgt19:AC090533 - 161.50 90.35 8.5e+03 241602 ! AC090533 Mus musculus strain  
gb\_hgt1:AC006842 - 161.50 88.46 1.1e+04 299015 ! AC006842 Caenorhabditis eleg  
gb\_in1:AE003433 - 161.50 88.33 1.1e+04 303411 ! AE003433 Drosophila melanoga  
gb\_pr10:HUM1GAA - 161.50 88.02 1.1e+04 314205 ! M36067 Human DNA ligase I mRNa  
gb\_pat2:AX072942 + 161.00 128.60 63.02 3083 ! AX072942 Sequence 53 from Pate  
gb\_pl3:D64087 + 161.00 127.67 70.95 3422 ! AX072942 Sequence 53 from Pate  
gb\_ov:GGBRG1PRT + 161.00 126.22 85.51 4033 ! D64087 Daucus carota mRNA for  
gb\_in1:AF036760 + 161.00 124.50 106.52 4893 ! X91637 G.gallus mRNA for BRG1  
gb\_in2:AF009623 + 161.00 123.30 124.35 5607 ! AF036760 Rattus norvegicus BRG  
gb\_in2:AF013277 + 161.00 122.65 135.06 6030 ! AF009623 Parascaris univalens  
gb\_in2:AF013277 + 161.00 122.11 144.82 6412 ! AF013277 Bombyx mori topoisome

seq\_name: gb\_pr8:HS26H231

seq\_documentation\_block:

LOCUS HS26H231 7444 bp mRNA PRI 25-JAN-2000  
DEFINITION Novel human gene mapping to chromosome 13.  
ACCESSION AL137201  
VERSION AL137201.1 GI:6759511

KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 7444)

REFERENCE  
AUTHORS Rhodes,S. and Huckle,E.

TITLE Direct Submission

JOURNAL Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk  
COMMENT This cDNA sequence was assembled from public domain ESTs and single  
pass sequencing reads from expressed DNA templates, aligned to the  
genomic DNA sequence from the bacterial clones 26H23 (284467),  
26P19 (275889) and 49U10 (284572).

The EST sequences listed match this sequence with an identity of at  
least 95% between the coordinates shown.  
Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene  
Sanger Centre name : 26H23.Cl3.1.

FEATURES

source

1. .7444  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13q12.3"  
1. .107  
/number=1  
108. .234  
/number=2  
127. .4470  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB69911.1"  
/db\_xref="GI:6759512"

exon

exon

CDS

/translation="MAHSKTRTNDGKITPPGVKEISDKISKEEMVRLKMKVKTMD  
MDDSEEEKLYLNALHSLASDFELKHPDKDVLVAVACCLADIFRIYAPAPVTS  
LKDFEFTQLKGLDTSKSPQNRYYLENTAWKSYNICELEDSNEIFQLYRT  
LFSYINGHNGKVMHVRDLMSSICEGDTSVQELDTVLVNLVPAHKNKQAYDIA  
KALLRTAQAEPTITNFQVLMGKTSISDLSEHFVDLILEYNDISLHVSLEPQ  
LEFKLSNDNEERLVYLLAKMFAGKDSLQNKPLQCYLGRFNDIHWPLRECV  
KFASHCLMNHDPDLAKDLEYLKVRSHPDEAIRDHIVTSVTAAKDILLVNDHLLNF  
VRERTLDKRWVRKEAMGLQIYKVYALQSAAGKDAKOAIWIKDKLLHYQNSID  
DRLLVERIPAQYVWPHNLETTERMKCLYLYAYLIDLNAVKALNEMKQNLRRQVND  
LLDLIKQPTDASVKAIIFSVMVITRNLPDPGKAQDFMKAFTQVLEDDERIKOLEVL  
VSPSCQKAEQVREITTKLGNPKPTNPFLEMIKFLERAPIRVHIDTESIALIKQ  
VNKSIDGTADEDEGVTDOAIRAGLELLKVLSTFHPISFHSAAETESLACLKMDDE  
KVAEAAQIFKNTGSKITEEDPHIRSAALLPVLLHKKSGKPPQAKYAIHCHIAIFSSK  
ETOPAQIFPELHSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSIVATFVKDIL  
MNDRLGPKTTKLWPDPEVSPETWYKIQAIKMWVRLKGNKNHNSKSGTSLRLTIT  
ILHSDGLTQGGKISKPSRSLRSLAAGSAIVKLAQEPYCEIITILEQYQCALAINDE  
CYQVQVFAQKLHGLSRLLPLEYMAICALCAKADPKVRRHARQCLVKNINVRREY  
LKQAAVSEKLLSLLPEYVPTTHLAHDPDYVKVODIEQLKDVRECLFWLEILMA  
KNNSHAFIRKNVNTKQVNTKQVNTKQVNTKQVNTKQVNTKQVNTKQVNTKQVNT  
KDPVLPARFTQDPKNSNTKNVLPPEMKSEFTPGPKPTNVLGAVNKPLSSAGKOSQ  
TKSSRMETVSNASSSSPSPGRIGKRLDSEMDHSENYDTWSSPLPGKSKDRRDS

DLVRSELEKPRGRKKTPTVTEQEBKLGMDLTKLVQEQPKGQSRKRKRGHTASDSDEQ  
QWPEERKRLKILENEDEQNSPKKGRGRPPKPLGGGTKEEPTMTKTSKKGKKKSG  
PPAPEEEEEERQSGNTEQKSKQHRVSRRAQQAESPSSAIESQTSTPQKGRGRP  
KTPSPSPQKKNVVRGSRKQATKENDSEEDVDFQGSPPVDDIPQETEEEEVSTVN  
VRRRSKAKRRR"  
235. .438  
/number=3  
439. .525  
/number=4  
526. .623  
/number=5  
624. .750  
/number=6  
751. .831  
/number=7  
832. .972  
/number=8  
973. .1088  
/number=9  
1089. .1183  
/number=10  
1184. .1329  
/number=11  
1330. .1481  
/number=12  
1482. .1595  
/number=13  
1596. .1677  
/number=14  
1678. .1726  
/number=15  
1727. .1866  
/number=16  
1866. .2046  
/note="matches EST AA687846 from clone IMAGE:1219676"  
1866. 1941  
/note="matches EST AA687776 from clone IMAGE:1219628"  
/note="matches EST  
1867. .1982  
/number=17  
1983. .2088  
/number=18  
2089. .2249  
/number=19  
2250. .2373  
/number=20  
2374. .2532  
/number=21  
2533. .2601  
/number=22  
2602. .2738  
/number=23  
2739. .2862  
/number=24  
2863. .3067  
/number=25  
complement(2912. .3302)  
/note="matches EST AA630616 from clone 856164"  
3068. .3182  
/number=26  
3183. .3315  
/number=27  
3302. .3572  
/note="matches EST AA333118"  
3316. .3435  
/number=28  
3439. .3818  
/note="matches EST AA282190 from clone IMAGE:704631"  
3436. .3498  
/number=29  
3499. .3644  
/number=30  
complement(3586. .4065)  
/note="matches EST AL655429 from clone IMAGE:2239331"



gb_pr10: HUMTRALL1	+ 164.50	117.79	251.88	14255	I04731 Homo sapiens translocat	gb_pr10: CLITOP04	+ 162.50	124.56	105.75	5557	Y16595 Cricetulus longicaud
gb_ov: GGRYR3	+ 164.50	117.29	268.69	15089	X95267 G.gallus mRNA for ryan	gb_ov: XELVTC	+ 162.50	124.46	107.07	5618	M18061 xenopus laevis vitel
gb_pi4: SPBPI6F5	- 164.50	112.54	494.25	25799	AL441603 S.pombe chromosome I	gb_in2: AVT0C068	+ 162.50	123.88	115.44	6013	D209068 Drosophila melanog
gb_hg1: AC008291	- 164.50	111.00	602.15	30696	AC008291 Drosophila melanogas	gb_in2: AFI16868	+ 162.50	120.46	178.90	8827	AFJ16868 Caenorhabditis ele
gb_p11: AB025628	- 164.50	102.50	1.8e+03	80117	AB025628 Arabidopsis thaliana	gb_in3: CEL243181	+ 162.50	120.45	179.25	8842	AJ243181 Caenorhabditis ele
gb_pr5: AF336797	- 164.50	96.19	4.0e+03	163217	AF336797 Homo sapiens small	gb_in3: CEL243179	+ 162.50	120.43	179.57	8856	AJ243179 Caenorhabditis ele
gb_hg16: AC073824	- 164.50	95.28	4.5e+03	180944	AC073824 Mus musculus clone	gb_in3: CEL243182	+ 162.50	120.40	180.33	8889	AJ243182 Caenorhabditis ele
gb_hg14: AC068066	- 164.50	95.19	4.6e+03	1802747	AC068066 Mus musculus chromo	gb_in3: CEL243180	+ 162.50	120.37	181.09	8922	AJ243180 Caenorhabditis ele
gb_hg15: AC069141	- 164.50	95.14	4.6e+03	180873	AC069141 Mus musculus chromo	gb_b1: AE005000	+ 162.50	118.13	241.32	11487	AE005000 Halobacterium sp.
gb_hg16: AC084410	- 164.50	94.30	5.1e+03	202075	O084410 Mus musculus chromo	gb_p14: SPBC56F2	- 162.50	106.91	1.0e+03	40765	AL023288 S.pombe chromosom
gb_hg2: AC010538	- 164.50	92.37	5.8e+03	224376	AC010538 Homo sapiens chromo	gb_pl1: AC073944	+ 162.50	106.12	1.1e+03	45450	AC073944 Arabidopsis thali
gb_hg16: AC073733	- 164.50	92.79	6.2e+03	239598	AC073733 Mus musculus clone	gb_hg5: AC017345	+ 162.50	106.06	1.1e+03	48688	AC017345 Drosophila melano
gb_hg13: AC051621	- 164.50	91.24	7.6e+03	285479	AC051621 Mus musculus clone	gb_ba3: SSOLPZNO9	+ 162.50	102.38	1.8e+03	67965	AL512970 Sulfolobus solfat
gb_pat2: E00869	+ 164.00	135.27	26.77	1896	E00869 cDNA encoding subunit	gb_hg1: AC011826	- 162.50	102.24	1.9e+03	69048	AC011826 Homo sapiens chro
gb_pr5: AC025361	- 164.00	133.88	32.01	2219	AC025361 Homo sapiens cDNA; FI	gb_pt3: AFVF5E19	+ 162.50	101.16	2.1e+03	77976	AL391147 Arabidopsis thali
gb_om: OCNFPROT5M	+ 164.00	133.08	35.46	2428	Z47378 O.cuniculus mrna for ne	gb_in1: AC004115	- 162.50	100.92	2.2e+03	80101	AC004115 Drosophila melano
gb_ba2: AF335605	+ 164.00	130.08	52.11	3407	AF335605 Borrelia burgdorferi	gb_in1: AC0040416	- 162.50	99.02	2.8e+03	92928	AC0040416 Caenorhabditis br
gb_pr9: HSBHVCV	+ 164.00	129.66	54.95	3570	X51591 Human beta-MHC mrna for	gb_hg5: AC017491	- 162.50	95.83	4.2e+03	142257	AC017491 Drosophila melan
gb_pi4: USMTQ1	+ 164.00	129.25	57.93	3740	L32017 Ustilago maydis topoisol	gb_hgt17: AC079885	+ 162.50	95.29	4.5e+03	151212	AC079885 Rattus norvegicus
gb_pr1: AC037748	- 164.00	126.32	84.34	5205	AC037748 Homo sapiens mrna for	gb_in1: AC008371	+ 162.50	94.60	4.9e+03	163461	AC008371 Drosophila melan
gb_pr5: AF1313003	- 164.00	121.92	148.45	8561	AF1313003 Homo sapiens silencer	gb_hgt4: AC007469	+ 162.50	93.68	5.6e+03	181446	AC007469 Drosophila melan
gb_r02: RNMRV7	- 164.00	121.90	148.42	8575	AT0011713 Rattus norvegicus mRN	gb_in1: AC012727	+ 162.50	93.16	5.9e+03	193255	AC012727 Drosophila melan
gb_ba1: AE0001288	+ 164.00	120.42	179.77	10132	AE001288 Chlamydia trachomat	gb_r01: AC073938	- 162.50	92.57	6.4e+03	205547	AC073938 Mus musculus l1
gb_ba1: AE002410	+ 164.00	118.29	236.35	12891	AE002410 Neisseria meningitidis	gb_in2: AE003779	+ 162.50	91.50	7.3e+03	232064	AE003779 Drosophila melan
gb_r01: AF136789	+ 164.00	117.05	277.03	14825	AF136789 Rattus norvegicus mu	gb_in2: AE003589	- 162.50	89.15	9.9e+03	302473	AE003589 Drosophila chro
gb_r01: AF227534	+ 164.00	116.40	301.28	15961	AF227534 Rattus norvegicus 3 BAC	gb_in3: DRDRI1P	+ 162.0	134.48	29.62	1734	J04725 D.melanogaster chrom
gb_pr3: AC016943	+ 164.00	95.13	4.6e+03	175992	AC016943 Homo sapiens 3 BAC	gb_pr9: HSM800656	+ 162.00	131.16	45.37	2524	AF080144 Homo sapiens mRNA;
gb_pr6: AL354928	- 164.00	95.10	4.6e+03	176539	AL354928 Human DNA sequence	gb_ba2: AF305608	+ 162.00	128.50	63.82	3408	AF305608 Borrelia burgdorfe
gb_pr5: AL136526	- 164.00	94.94	4.7e+03	179804	AL136526 Human DNA sequence	gb_ba2: AF305610	+ 162.00	127.65	71.13	3749	AF305610 Borrelia burgdorfe
gb_hg95: AC016957	- 164.00	94.87	4.8e+03	181244	AC016957 Homo sapiens chromo	gb_pr1: AB046771	+ 162.00	127.47	72.85	3829	AB046771 Homo sapiens mRNA
gb_hg19: AC009552	- 164.00	94.84	4.8e+03	181906	AC009552 Homo sapiens chromo	gb_pr1: AB016091	+ 162.00	126.29	84.72	4373	M74099 Human displacement p
gb_hg1: AC007343	- 164.00	94.56	5.0e+03	187701	AC007343 Homo sapiens chromo	gb_pr10: HUMCDP	+ 162.00	124.46	107.13	5376	M74099 Human displacement p
gb_hg12: AC027682	- 164.00	94.56	5.0e+03	203889	AC027682 Homo sapiens chromo	gb_pr5: AFI11785	+ 162.00	123.60	119.64	5925	AF111785 Homo sapiens mvys1
gb_hg20: AL158075	- 164.00	93.83	5.4e+03	203889	AL158075 Homo sapiens chromo	gb_pr10: HUMIP3R3	+ 162.00	120.06	188.34	8833	U01062 Human type 3 insu
gb_hg24: CEV7682	- 164.00	91.37	7.5e+03	269082	Z92866 Caenorhabditis elegans	gb_ba1: AE001457	- 162.00	118.40	233.14	10658	AE001457 Helicobacter pylor
gb_in1: AE003495	+ 164.00	90.16	8.7e+03	308373	AE003495 Drosophila melanoga	gb_pi4: SCYLR106C	+ 162.00	114.83	368.21	15935	Z73278 S.cerevisiae chromo
gb_pat1: AX044029	+ 164.00	89.04	1.0e+04	349980	AX044029 Sequence 108 from H	gb_hg6: AC019925	- 162.00	108.64	814.80	32058	AC019925 Drosophila melano
gb_ov: AF144637	+ 163.50	134.56	29.33	1965	AF144637 Gallus gallus osteobla	gb_pi4: SPAC23A1	- 162.00	107.67	922.79	35769	AL021813 S.pombe chromosom
gb_ov: GGA277602	+ 163.50	134.56	29.33	1965	AF144637 Gallus gallus osteobla	gb_pi4: YSCL8004	- 162.00	106.60	1.1e+03	40340	U53876 Saccharomyces cerev
gb_pr7: BC002781	+ 163.50	133.69	32.78	2167	BC002781 Homo sapiens, Similar	gb_pl3: NCB268	- 162.00	100.93	2.2e+03	76499	AL355930 Neurospora crassa
gb_r01: BC003273	+ 163.50	133.49	33.66	2218	BC003273 Mus musculus, Clone	gb_pr4: AC026273	- 162.00	95.29	4.5e+03	144645	AC026273 Homo sapiens clo
gb_r02: S68108	+ 163.50	129.41	56.75	3513	S68108 brgl-brahma homolog (m	gb_in1: AC006170	- 162.00	94.07	5.3e+03	166105	AC024972 Homo sapiens clo
gb_ov: CHRSMCA	+ 163.50	128.03	67.80	4108	X04968 Chicken smooth muscle c	gb_in1: AC006170	- 162.00	93.94	5.4e+03	168471	AC006170 Drosophila melan
gb_r02: RNSYNP2	+ 163.50	126.70	80.36	4771	X08981 R.norvegicus mrna for s	gb_hgt1: AC027276	- 162.00	93.15	5.9e+03	184244	AC027276 Mus musculus clo
gb_in1: AC006657	- 163.50	124.41	107.85	6181	AC006657 Caenorhabditis elegans	gb_hgt1: AC023611	- 162.00	91.52	7.3e+03	221285	AC023611 Mus musculus clo
gb_pr1: AB007934	- 163.50	124.26	109.85	6282	AB007934 Homo sapiens mrna for	gb_in2: AE003767	- 162.00	91.37	7.5e+03	225155	AE003767 Drosophila melan
gb_p11: AB934154	+ 163.50	121.60	154.60	8486	AB934154 Chara corallina ccm m	gb_in1: AE003446	- 162.00	88.74	1.0e+04	303092	AE003446 Drosophila melan
gb_pat1: AB91752	+ 163.50	120.42	179.90	9697	AB91752 Sequence 8 from Patent	gb_r03: RADENTMAT	+ 161.50	129.93	53.11	2773	L11354 Rattus norvegicus ac
gb_in2: ANDDHBC	+ 163.50	116.60	293.50	14918	D01021 Anthocidaris crassipet	gb_r03: RADENTMAT	+ 161.50	128.25	55.86	3351	L11354 Rattus norvegicus ac
gb_in2: AF002180	+ 163.50	114.77	326.32	16377	AF002180 Plasmodium falciparu	gb_in4: PRAGP195	+ 161.50	124.23	110.31	5276	M37213 P.falciparum major m
gb_hg4: AC015359	- 163.50	114.55	381.64	18797	AC015359 Drosophila melanogas	gb_pr1: AC023950	+ 161.50	124.20	110.72	5293	AB032950 Homo sapiens mRNA
gb_hg2: AC009190	+ 163.50	108.72	806.74	36323	AC009190 Homo sapiens chromo	gb_pr9: HSM801713	+ 161.50	123.97	114.07	5434	X63692 H.sapiens mRNA for D
gb_pl3: NCB384	+ 163.50	103.92	1.5e+03	62431	AC059130 Neurospora crassa DN	gb_pr9: HSM801713	+ 161.50	123.91	114.91	5469	AL136745 Homo sapiens mRNA;
gb_hg18: AC087139	+ 163.50	100.57	2.3e+03	91061	AC087139 Mus musculus clone R	gb_pr9: HSGRESTIN	+ 161.50	123.31	124.22	5857	AL136745 Homo sapiens mRNA;
gb_hg9: AC025473	- 163.50	99.87	2.5e+03	98572	AC025473 Homo sapiens chromo	gb_pat1: AX010195	+ 161.50	107.28	969.63	35734	X64838 H.sapiens mrna for r
gb_v11: AF081810	- 163.50	95.52	4.4e+03	161046	AF081810 Lymantiria dispar nu	gb_hg6: AC019974	+ 161.50	119.48	202.87	9019	AX010195 Sequence 1 from Pa
gb_hg9: AC0033611	- 163.50	92.71	6.3e+03	221285	AC0033611 Mus musculus clone	gb_in3: CF05P051	+ 161.50	107.03	1.0e+03	36759	Z81586 Caenorhabditis eleg
gb_in2: AE003686	- 163.50	92.68	6.3e+03	221888	AE003686 Drosophila melanoga	gb_hg4: AC013883	+ 161.50	105.17	1.3e+03	45371	AC013883 Drosophila melano
gb_sts2: CNS070ECB	- 163.00	139.53	15.50	1072	AL441585 T7 end of clone XBD0	gb_hg6: AC019768	+ 161.50	104.64	1.4e+03	48164	AC019768 Drosophila melano
gb_om: CFP180RRC	+ 163.00	125.17	97.82	5425	X87224 Canis familiaris mrna f	gb_hgt6: AC019768	+ 161.50	99.64	2.6e+03	84679	AC022571 Homo sapiens cion
gb_r01: AF110498	+ 163.00	124.49	106.67	5855	AF110498 Mesocricetus auratus	gb_hgt6: AC019768	+ 161.50	97.11	3.6e+03	112726	AC022571 Homo sapiens cion
gb_pr1: AB007923	+ 163.00	123.32	123.93	6081	AB007923 Homo sapiens mrna for	gb_hgt24: CEY21F11	+ 161.50	99.61	3.9e+03	122061	AL009107 Caenorhabditis e
gb_ba1: AE002514	+ 163.00	118.95	217.20	10947	AE002514 Neisseria meningitidis	gb_hgt1: AC008324	+ 161.50	96.21	4.0e+03	122061	AC008324 Drosophila melan
gb_in2: AF000580	+ 163.00	116.18	309.61	14955	AF000580 Dictyostelium discoi	gb_in1: AC005558	- 161.50	96.21	4.0e+03	124700	AC005558 Drosophila melan
gb_hg11: AC027405	+ 163.00	103.03	1.7e+03	66033	AC027405 Homo sapiens chromo	gb_hgt4: AC012615	- 161.50	94.60	4.9e+03	149515	AC012615 Homo sapiens chr
gb_hg15: AC073669	- 163.00	95.58	4.3e+03	152368	AC073669 Mus musculus clone	gb_hgt3: AC012431	+ 161.50	94.09	5.3e+03	158430	AC012431 Homo sapiens clo
gb_hg10: AC026106	- 163.00	94.94	4.7e+03	164485	AC026106 Homo sapiens chromo	gb_hgt5: AC069284	+ 161.50	94.05	5.3e+03	159234	AC069284 Homo sapiens chr
gb_hg11: AC026319	+ 163.00	93.47	5.7e+03	194182	AC026319 Homo sapiens chromo	gb_hgt8: AC022248	- 161.50	93.97	5.3e+03	160584	AC022248 Homo sapiens clo
gb_hg3: AC012474	+ 163.00	93.18	5.9e+03	200579	AC012474 Homo sapiens chromo	gb_pr1: AC025540	- 161.50	93.95	5.4e+03	160896	AC025540 Homo sapiens clo
gb_pat2: E02461	+ 162.50	134.07	31.22	1899	E02461 cDNA encoding glycine	gb_hgt8: AC023270	- 161.50	93.34	5.8e+03	172437	AC023270 Homo sapiens chr
gb_pl3: GMGLYR1	+ 162.50	134.07	31.22	1899	X02626 Soybean mrna for A5A4B	gb_in1: AC009392	- 161.50	93.00	6.1e+03	179124	AC009392 Drosophila melan
gb_in3: CNU00168	+ 162.50	128.48	63.94	3569	U20168 Caenorhabditis elegans	gb_hgt19: AC090442	+ 161.50	92.88	6.2e+03	181715	AC090442 Homo sapiens chr
gb_r01: CEUOTOPIS1	+ 162.50	124.59	105.29	5536	L04607 Cricetulus griseus DNA	gb_v11: AF1232689	+ 161.50	90.79	8.2e+03	229896	AF1232689 Rat cytomegalovir
						gb_hgt17: AC079427	+ 161.50	90.59	8.2e+03	235188	AC079427 Mus musculus chr



gb_inl:AE003489	+	168.00	93.48	5.7e+03	302915	!	AE003489 Drosophila melanog
gb_pr10:HUMSRP1570	-	167.50	137.23	20.82	2076	!	LI4076 Human pre-mRNA splicing
gb_pr9:HSAC001050	-	167.50	133.46	33.77	3177	!	AC001070 Homo sapiens (Subclon
gb_pl3:SC34225	+	167.50	129.15	58.67	5166	!	D84225 Arabidopsis thaliana fr
gb_ov:D89991	+	167.50	127.92	68.77	5941	!	D89991 Cyprinus carpio mRNA f
gb_pr5:AF111784	+	167.50	127.89	68.97	5956	!	AF111784 Homo sapiens myosin I
gb_ov:XELTANABIN	+	167.50	126.44	83.11	7019	!	M93387 Xenopus laevis tanabin
gb_hg6:AC020364	+	167.50	124.74	103.31	8500	!	AC020364 Drosophila melanogast
gb_pr8:HS1340191	+	167.50	122.34	140.55	11145	!	AF096858 Novel human gene map
gb_r01:AF150755	+	167.50	118.43	232.15	17333	!	AF150755 Mus musculus microtu
gb_ba1:AE000561	-	167.50	117.79	251.97	18629	!	AE000561 Helicobacter pylori
gb_hg14:AC068192	-	167.50	110.29	659.02	43417	!	AC068192 Homo sapiens clone F
gb_pr1:AB0065702	-	167.50	104.21	1.4e+03	86263	!	AB006702 Arabidopsis thaliana
gb_hg21:AL355394	-	167.50	104.15	1.4e+03	86897	!	AL355394 Human DNA sequence
gb_hg11:AC009257	-	167.50	98.90	2.8e+03	157095	!	AL355994 Homo sapiens chromo
gb_inl:AC009257	-	167.50	97.37	3.5e+03	186827	!	AC026478 Mus musculus chromo
gb_hg16:AC074307	-	167.50	95.98	4.1e+03	218565	!	AC092527 Drosophila melanog
gb_hg17:AC079431	-	167.50	95.90	4.2e+03	220469	!	AC074307 Mus musculus chromo
gb_pr10:AB046029	+	167.50	93.15	5.9e+03	300695	!	AC079431 Mus musculus chromo
gb_ba2:AF305602	+	167.00	136.76	22.13	2095	!	AB046029 Macaca fascicularis E
gb_ba2:AF305600	+	167.00	132.04	40.54	3569	!	AF305600 Borrelia burgdorferi
gb_ba2:AF305604	+	167.00	132.04	40.54	3569	!	AF305602 Borrelia burgdorferi
gb_ba2:AF305606	+	167.00	132.04	40.54	3569	!	AF305604 Borrelia burgdorferi
gb_ba2:AF305609	+	167.00	132.04	40.54	3569	!	AF305606 Borrelia burgdorferi
gb_pr10:HSU40847	+	167.00	131.64	42.64	3731	!	AF305609 Borrelia burgdorferi
gb_ov:AF017250	+	167.00	129.43	56.62	4789	!	U40847 Human Treacher Collins
gb_ba1:AE001131	+	167.00	129.34	57.31	4840	!	U6366 Human Treacher Collins
gb_ba1:AE001153	+	167.00	128.19	66.42	5511	!	AF017250 Oreochromis aureus vi
gb_v12:HHU43400	-	167.00	121.15	163.70	12190	!	AE001131 Borrelia burgdorferi
gb_hg5:AF037128	+	167.00	118.17	240.00	17070	!	AE001153 Borrelia burgdorferi
gb_hg10:AC025326	-	167.00	106.77	1.0e+03	61843	!	AL353822 Neurospora crassa DN
gb_hg3:AC012474	-	167.00	99.22	2.7e+03	144861	!	U43400 Human herpesvirus-7 C
gb_hg16:AC073809	+	167.00	98.93	2.8e+03	149845	!	AC016932 Homo sapiens chromo
gb_ba2:BUHCY54	+	166.50	98.74	2.9e+03	153080	!	AF037218 Human herpesvirus 7
gb_in4:PEGP195	+	167.00	98.05	3.2e+03	165490	!	AC025326 Homo sapiens chromo
gb_pr5:AF111783	+	166.50	97.97	3.2e+03	166991	!	AC090884 Homo sapiens chromo
gb_pr3:SCU03673	+	167.00	96.34	3.9e+03	200379	!	AC012474 Homo sapiens chromo
gb_pat2:E26748	+	166.50	93.43	5.7e+03	278652	!	AC073809 Mus musculus clone
gb_pat2:AX067150	+	166.50	93.15	27.27	2408	!	AK033416 Homo sapiens cDNA FLJ
gb_in3:CELSPC3A	-	166.50	129.52	55.97	4534	!	M90644 Buchnera aphidicola DNA
gb_in3:CELSPC3A	-	166.50	129.29	57.67	4655	!	X03831 P.falciparum (CAMP stra
gb_in3:CELSPC3A	-	166.50	127.74	70.30	5541	!	AF111783 Homo sapiens myosin I
gb_in3:CELSPC3A	-	166.50	127.01	77.19	6016	!	AC03673 Saccharomyces cerevisia
gb_in3:CELSPC3A	-	166.50	126.86	78.78	6125	!	E26748 Structure and function
gb_in3:CELSPC3A	-	166.50	126.83	79.07	6145	!	AX067150 Sequence 2 from Paten
gb_in3:CELSPC3A	-	166.50	125.95	88.45	6782	!	AX067151 Sequence 3 from Paten
gb_in3:CELSPC3A	-	166.50	124.57	105.68	7932	!	M85149 Caenorhabditis elegans
gb_in3:CELSPC3A	-	166.50	123.04	128.54	9424	!	AE000770 Aquifex aeolicus sec
gb_in3:CELSPC3A	-	166.50	119.60	199.90	13900	!	Z47814 S.cerevisiae chromosom
gb_in3:CELSPC3A	-	166.50	119.02	215.09	14825	!	U10399 Saccharomyces cerevisi
gb_in3:CELSPC3A	-	166.50	113.35	445.23	28124	!	X95966 S.cerevisiae 35.4kb DN
gb_in3:CELSPC3A	-	166.50	111.30	579.23	35452	!	AC027282 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	109.62	718.21	42839	!	AF078790 Caenorhabditis eleg
gb_in3:CELSPC3A	-	166.50	109.41	738.59	43907	!	U33050 Saccharomyces cerevisi
gb_in3:CELSPC3A	-	166.50	105.40	1.2e+03	69023	!	AC06550 Arabidopsis thaliana
gb_in3:CELSPC3A	-	166.50	104.28	1.4e+03	78341	!	AL35798 Human DNA sequence
gb_in3:CELSPC3A	-	166.50	101.83	2.0e+03	103209	!	AC08632 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	100.99	2.2e+03	113522	!	AC069401 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	98.57	3.0e+03	149125	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	97.75	3.3e+03	163567	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	96.35	3.9e+03	191575	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.00	142.34	10.81	1020	!	AL116784 Botrytis cinerea stra
gb_in3:CELSPC3A	-	166.00	134.67	28.92	2425	!	AF17292 Homo sapiens genethon
gb_in3:CELSPC3A	-	166.00	132.63	37.55	3052	!	AF154023 Streptococcus pneumo
gb_in3:CELSPC3A	-	166.00	132.48	38.31	3106	!	AF19085 Homo sapiens bromodon
gb_in3:CELSPC3A	-	166.00	128.28	65.67	4991	!	AF127481 Homo sapiens non-ocod
gb_in3:CELSPC3A	-	166.00	127.77	70.05	5283	!	AB040942 Homo sapiens mRNA for
gb_in3:CELSPC3A	-	166.00	127.46	72.89	5283	!	X05640 Mouse NF-M gene for mid
gb_in3:CELSPC3A	-	166.00	126.71	80.22	5471	!	D89547 Chicken mRNA for vitell
gb_in3:CELSPC3A	-	166.00	126.66	80.75	5987	!	AF059399 Bos taurus MyHC-2x m
gb_in3:CELSPC3A	-	166.00	126.65	80.91	5997	!	AF212149 Oryctolagus cuniculu
gb_in3:CELSPC3A	-	166.00	126.59	81.53	6038	!	AF212148 Oryctolagus cuniculu
gb_inl:AE003489	+	168.00	93.48	5.7e+03	302915	!	AE003489 Drosophila melanog
gb_pr10:HUMSRP1570	-	167.50	137.23	20.82	2076	!	LI4076 Human pre-mRNA splicing
gb_pr9:HSAC001050	-	167.50	133.46	33.77	3177	!	AC001070 Homo sapiens (Subclon
gb_pl3:SC34225	+	167.50	129.15	58.67	5166	!	D84225 Arabidopsis thaliana fr
gb_ov:D89991	+	167.50	127.92	68.77	5941	!	D89991 Cyprinus carpio mRNA f
gb_pr5:AF111784	+	167.50	127.89	68.97	5956	!	AF111784 Homo sapiens myosin I
gb_ov:XELTANABIN	+	167.50	126.44	83.11	7019	!	M93387 Xenopus laevis tanabin
gb_hg6:AC020364	+	167.50	124.74	103.31	8500	!	AC020364 Drosophila melanogast
gb_pr8:HS1340191	+	167.50	122.34	140.55	11145	!	AF096858 Novel human gene map
gb_r01:AF150755	+	167.50	118.43	232.15	17333	!	AF150755 Mus musculus microtu
gb_ba1:AE000561	-	167.50	117.79	251.97	18629	!	AE000561 Helicobacter pylori
gb_hg14:AC068192	-	167.50	110.29	659.02	43417	!	AC068192 Homo sapiens clone F
gb_pr1:AB0065702	-	167.50	104.21	1.4e+03	86263	!	AB006702 Arabidopsis thaliana
gb_hg21:AL355394	-	167.50	104.15	1.4e+03	86897	!	AL355394 Human DNA sequence
gb_hg11:AC009257	-	167.50	98.90	2.8e+03	157095	!	AL355994 Homo sapiens chromo
gb_inl:AC009257	-	167.50	97.37	3.5e+03	186827	!	AC026478 Mus musculus chromo
gb_hg16:AC074307	-	167.50	95.98	4.1e+03	218565	!	AC092527 Drosophila melanog
gb_hg17:AC079431	-	167.50	95.90	4.2e+03	220469	!	AC074307 Mus musculus chromo
gb_pr10:AB046029	+	167.50	93.15	5.9e+03	300695	!	AC079431 Mus musculus chromo
gb_ba2:AF305602	+	167.00	136.76	22.13	2095	!	AB046029 Macaca fascicularis E
gb_ba2:AF305600	+	167.00	132.04	40.54	3569	!	AF305600 Borrelia burgdorferi
gb_ba2:AF305604	+	167.00	132.04	40.54	3569	!	AF305602 Borrelia burgdorferi
gb_ba2:AF305606	+	167.00	132.04	40.54	3569	!	AF305604 Borrelia burgdorferi
gb_ba2:AF305609	+	167.00	132.04	40.54	3569	!	AF305606 Borrelia burgdorferi
gb_pr10:HSU40847	+	167.00	131.64	42.64	3731	!	AF305609 Borrelia burgdorferi
gb_ov:AF017250	+	167.00	129.43	56.62	4789	!	U40847 Human Treacher Collins
gb_ba1:AE001131	+	167.00	129.34	57.31	4840	!	U6366 Human Treacher Collins
gb_ba1:AE001153	+	167.00	128.19	66.42	5511	!	AF017250 Oreochromis aureus vi
gb_v12:HHU43400	-	167.00	121.15	163.70	12190	!	AE001131 Borrelia burgdorferi
gb_hg5:AF037128	+	167.00	118.17	240.00	17070	!	AE001153 Borrelia burgdorferi
gb_hg10:AC025326	-	167.00	106.77	1.0e+03	61843	!	AL353822 Neurospora crassa DN
gb_hg3:AC012474	-	167.00	99.22	2.7e+03	144861	!	U43400 Human herpesvirus-7 C
gb_hg16:AC073809	+	167.00	98.93	2.8e+03	149845	!	AC016932 Homo sapiens chromo
gb_ba2:BUHCY54	+	166.50	98.74	2.9e+03	153080	!	AF037218 Human herpesvirus 7
gb_in4:PEGP195	+	167.00	98.05	3.2e+03	165490	!	AC025326 Homo sapiens chromo
gb_pr5:AF111783	+	166.50	97.97	3.2e+03	166991	!	AC090884 Homo sapiens chromo
gb_pr3:SCU03673	+	167.00	96.34	3.9e+03	200379	!	AC012474 Homo sapiens chromo
gb_pat2:E26748	+	166.50	93.43	5.7e+03	278652	!	AC073809 Mus musculus clone
gb_pat2:AX067150	+	166.50	93.15	27.27	2408	!	AK033416 Homo sapiens cDNA FLJ
gb_in3:CELSPC3A	-	166.50	129.52	55.97	4534	!	M90644 Buchnera aphidicola DNA
gb_in3:CELSPC3A	-	166.50	129.29	57.67	4655	!	X03831 P.falciparum (CAMP stra
gb_in3:CELSPC3A	-	166.50	127.74	70.30	5541	!	AF111783 Homo sapiens myosin I
gb_in3:CELSPC3A	-	166.50	127.01	77.19	6016	!	AC03673 Saccharomyces cerevisia
gb_in3:CELSPC3A	-	166.50	126.86	78.78	6125	!	E26748 Structure and function
gb_in3:CELSPC3A	-	166.50	126.83	79.07	6145	!	AX067150 Sequence 2 from Paten
gb_in3:CELSPC3A	-	166.50	125.95	88.45	6782	!	AX067151 Sequence 3 from Paten
gb_in3:CELSPC3A	-	166.50	124.57	105.68	7932	!	M85149 Caenorhabditis elegans
gb_in3:CELSPC3A	-	166.50	123.04	128.54	9424	!	AE000770 Aquifex aeolicus sec
gb_in3:CELSPC3A	-	166.50	119.60	199.90	13900	!	Z47814 S.cerevisiae chromosom
gb_in3:CELSPC3A	-	166.50	119.02	215.09	14825	!	U10399 Saccharomyces cerevisi
gb_in3:CELSPC3A	-	166.50	113.35	445.23	28124	!	X95966 S.cerevisiae 35.4kb DN
gb_in3:CELSPC3A	-	166.50	111.30	579.23	35452	!	AC027282 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	109.62	718.21	42839	!	AF078790 Caenorhabditis eleg
gb_in3:CELSPC3A	-	166.50	109.41	738.59	43907	!	U33050 Saccharomyces cerevisi
gb_in3:CELSPC3A	-	166.50	105.40	1.2e+03	69023	!	AC06550 Arabidopsis thaliana
gb_in3:CELSPC3A	-	166.50	104.28	1.4e+03	78341	!	AL35798 Human DNA sequence
gb_in3:CELSPC3A	-	166.50	101.83	2.0e+03	103209	!	AC08632 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	100.99	2.2e+03	113522	!	AC069401 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	98.57	3.0e+03	149125	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	97.75	3.3e+03	163567	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.50	96.35	3.9e+03	191575	!	AL365325 Homo sapiens chromo
gb_in3:CELSPC3A	-	166.00	142.34	10.81	1020	!	AL116784 Botrytis cinerea stra
gb_in3:CELSPC3A	-	166.00	134.67	28.92	2425	!	AF17292 Homo sapiens genethon
gb_in3:CELSPC3A	-</						



gb_hlg6:AC020202	-	172.00	104.94	1.3e+03	118591	! AC020202 Drosophila melanoga	gb_prio:HUMAF4Y	+	170.00	125.83	89.82	9390	! LI3773 Human AF-4 mRNA, com
gb_in1:AC007837	-	172.00	104.57	1.4e+03	123647	! AC007837 Drosophila melanoga	gb_pat2:143747	+	170.00	125.83	89.83	9391	! I43747 Sequence 25 from pat
gb_in2:AC009218	-	172.00	102.47	1.8e+03	126753	! AC009218 Drosophila melanoga	gb_in3:DM2245406	+	170.00	113.66	428.03	37107	! A124506 Drosophila melano
gb_hlg11:AC026903	-	172.00	102.34	1.8e+03	153179	! AC026903 Homo sapiens chromo	gb_hlg5:AC017329	+	170.00	105.24	1.3e+03	95361	! AC017329 Drosophila melano
gb_in29:AC009551	-	172.00	99.37	2.7e+03	222079	! AC009551 Homo sapiens chromo	gb_prio:AC004903	+	170.00	100.34	1.6e+03	118869	! AC004903 Homo sapiens PAC
gb_hlg19:AC003791	-	172.00	97.97	3.2e+03	260686	! AC003791 Drosophila melanoga	gb_prio:AC011456	+	170.00	100.92	2.2e+03	156357	! AC011456 Homo sapiens chr
gb_ov:HSX8011327	+	171.50	134.74	28.65	3926	! AL133161 Homo sapiens mRNA; cl	gb_in1:AC007830	+	170.00	97.65	3.3e+03	226168	! AC007830 Mus musculus 10
gb_ov:XLAS21130	+	171.50	130.70	48.13	6198	! J251130 Xenopus laevis mRNA f	gb_in1:AC007830	+	170.00	97.65	3.3e+03	226576	! AL590447 chromosome VII o
gb_ov:AC022655	+	171.50	128.65	62.63	7814	! AF022655 Homo sapiens cep250 c	gb_ov:RATNEM	+	169.50	136.43	23.09	2717	! M18628 Rat NF-M middle mole
gb_in2:AR093136	+	171.50	127.82	69.60	8575	! AF093136 Toxoplasma gondii DNA	gb_ov:AC0269340	+	169.50	132.93	31.81	3602	! AF269340 Staphylococcus epi
gb_ov:GGY19187	+	171.50	122.45	138.55	15717	! Y19187 Gallus gallus mRNA for	gb_in2:AF183401	+	169.50	132.64	37.51	4165	! AF183401 Caenorhabditis ele
gb_ov:SPBC336	+	171.50	114.29	394.92	39512	! AL121815 S.pombe chromosome 1	gb_pat1:AS92450	+	169.50	131.16	45.33	4920	! A92450 Sequence 1 from date
gb_in3:CEUK11C4	+	171.50	113.54	434.52	42978	! U64854 Caenorhabditis elegans	gb_prio:HUMHNF2B	+	169.50	130.67	48.27	5200	! D26156 Human mRNA for trans
gb_ov:AC000547	+	171.50	104.21	1.4e+03	123288	! AC000547 Homo sapiens genom	gb_prio:HSU29175	+	169.50	130.59	48.77	5247	! U29175 Human transcrip
gb_hlg5:AC017582	+	171.50	103.16	1.6e+03	133685	! AC017582 Drosophila melanoga	gb_prio:AB014886	+	169.50	130.11	51.92	5544	! AB014886 Emericella nidulan
gb_ov:AC019140	+	171.50	101.21	2.1e+03	172868	! AC019140 Homo sapiens clone	gb_ov:AF126834	+	169.50	129.59	55.44	5874	! AF126834 Mus musculus 8prip
gb_hlg7:AC020796	+	171.50	98.52	3.0e+03	234125	! AC020796 Mus musculus clone	gb_ov:PFANT195	+	169.50	129.53	59.91	5917	! X02919 Plasmodium falciparu
gb_in1:AE003536	+	171.50	95.18	4.6e+03	341319	! AE003536 Drosophila melanoga	gb_ov:AC022023	+	169.50	128.99	59.95	6292	! AC022023 Bos taurus mra fo
gb_in3:GLSR2	+	171.00	132.40	38.69	4892	! X79815 G.lambalia SR2 gene, 671	gb_in2:PRGP195A	+	169.50	127.99	68.09	7038	! X15063 Plasmodium falciparu
gb_ov:CHKVTC	+	171.00	130.90	46.87	5791	! M18060 Chicken vitellogenin ge	gb_in2:AC298215	+	169.50	127.85	69.35	7153	! AF298215 Drosophila melano
gb_ov:AC02322	+	171.00	130.75	47.82	5984	! AB023222 Homo sapiens mRNA for	gb_prio:HS85D21	+	169.50	124.27	109.70	10709	! AL049784 Novel human gene
gb_ov:NC0243769	+	171.00	130.62	48.63	5982	! AJ243769 Notothenia coriiceps	gb_prio:AF020713	+	169.50	101.86	1.9e+03	134416	! AF020713 Bacteriophage SP
gb_p13:CAROP2	+	171.00	130.17	51.48	6290	! Y10377 C.albicans TOP2 gene, 5	gb_prio:AF020713	+	169.50	101.86	2.4e+03	161573	! AF020713 Human herpesviru
gb_prio:AF201422	+	171.00	128.28	65.64	7789	! AF201422 Homo sapiens splicing	gb_in1:AC009183	+	169.50	98.70	2.9e+03	192055	! AC009183 Drosophila melan
gb_pat1:AG1387	+	171.00	128.02	67.87	8022	! A61387 Sequence 3 from Patent	gb_hlg9:AC023833	+	169.50	98.25	3.1e+03	202083	! AC023833 Mus musculus chr
gb_p14:SPRAD3GEN	+	171.00	128.02	67.87	8022	! Y09076 S.pombe RAD3 gene, 4719	gb_hlg6:BSU00012	+	169.50	97.75	3.3e+03	213680	! Z99115 Bacillus subtilis
gb_p14:SPU76307	+	171.00	128.02	67.87	8022	! U76307 Schizosaccharomyces pom	gb_hlg16:AC074307	+	169.50	97.48	3.4e+03	220469	! AC074307 Mus musculus chr
gb_prio:AB016092	+	171.00	126.97	77.61	9027	! AB016092 Homo sapiens mRNA for	gb_in2:AE003685	+	169.50	97.30	3.5e+03	225038	! AE003685 Drosophila melan
gb_ov:AE000600	+	171.00	124.10	112.24	12489	! AE000600 Helicobacter pylori	gb_hlg6:AC017387	+	169.50	102.74	102.74	9668	! AC017387 Drosophila melano
gb_ov:U39701	+	171.00	122.36	140.21	15191	! U39701 Mycobacter genitalium	gb_hlg6:AC017670	+	169.50	113.21	453.40	35708	! AC017670 Drosophila melano
gb_p14:SPBC216	+	171.00	118.24	140.21	15191	! U39701 Mycobacter genitalium	gb_prio:AF130343	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_p14:AC004493	+	171.00	113.01	265.45	25311	! AL049558 S.pombe chromosome 1	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_p14:AC004493	+	171.00	113.01	265.45	25311	! AL049558 S.pombe chromosome 1	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg21:AL359955	+	171.00	101.35	2.1e+03	162782	! AL359955 Homo sapiens chromo	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AC00719	+	171.00	99.68	2.6e+03	196424	! AP000719 Homo sapiens genom	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg18:AC087566	+	171.00	98.21	3.1e+03	231912	! AL459933 Homo sapiens chromo	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg22:AL445993	+	171.00	97.97	3.2e+03	238351	! AL445993 Homo sapiens chromo	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in2:AC003824	+	171.00	97.58	3.4e+03	249001	! AE003824 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg7:AC020870	+	171.00	97.32	3.5e+03	256373	! AC020870 Mus musculus clone	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg7:BC003741	+	170.50	138.89	16.83	22649	! BC003741 Mus musculus. Similar	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:HUMHMCARA	+	170.50	135.26	26.81	3388	! D10667 Homo sapiens mRNA for S	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in4:PFPL90	+	170.50	131.33	44.41	5282	! X03371 P.falciparum gp190 (MS	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_p14:SPAC5F8	+	170.50	114.06	40.73	37093	! Z69728 S.pombe chromosome 1 C	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AC005937	+	170.50	111.90	536.41	47323	! AC005937 Homo sapiens clone t	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg4:AC013847	+	170.50	111.50	564.42	49491	! AC013847 Drosophila melanogas	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg1:AC005979	+	170.50	108.55	824.51	69084	! AC005979 Drosophila melanogas	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg6:AC018038	+	170.50	106.81	1.0e+03	84011	! AC018038 Drosophila melanogas	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_p11:AC079041	+	170.50	103.70	1.5e+03	119420	! AC079041 Arabidopsis thalian	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg6:AC020329	+	170.50	101.39	2.2e+03	154895	! AC020329 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AC084730	+	170.50	100.73	2.2e+03	166929	! AC084730 Papio hamadryas cl	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in1:AC008916	+	170.50	100.61	2.3e+03	169271	! AC008916 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg7:AC020858	+	170.50	99.97	2.5e+03	181955	! AC020858 Mus musculus clone	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AB023048	+	170.50	99.76	2.5e+03	186222	! AC020858 Mus musculus clone	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg1:AB023048	+	170.50	99.46	2.7e+03	192650	! AB023048 Homo sapiens genom	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg1:AC084071	+	170.50	99.31	2.8e+03	195906	! AC084071 Homo sapiens clone	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AF000511	+	170.50	99.13	2.8e+03	200000	! AP000511 Homo sapiens genom	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in1:AC008237	+	170.50	99.12	2.8e+03	200216	! AC008237 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg22:AL599737	+	170.50	99.00	2.8e+03	202939	! AL589737 Mus musculus chromo	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg18:AC097227	+	170.50	98.88	2.8e+03	205691	! AC087227 Mus musculus chromo	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in2:AE003748	+	170.50	96.92	3.7e+03	256608	! AE003748 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in1:AE003460	+	170.50	95.65	4.3e+03	279907	! AE003460 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in1:AE003438	+	170.50	95.55	4.4e+03	299537	! AE003438 Drosophila melanoga	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in2:AF162774	+	170.00	135.89	24.73	3018	! AF162774 Drosophila melanogast	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_ov:RNNFMR	+	170.00	135.74	25.22	3070	! Z12152 R.norvegicus mRNA for n	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:RATNP140A	+	170.00	134.30	30.30	3609	! M94287 Rattus norvegicus nucle	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:HUMSPRP	+	170.00	134.21	30.37	3647	! L25050 Human serine/proline-ri	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_ov:CFCS3VS	+	170.00	132.07	40.35	4643	! X99145 Canis familiaris mRNA f	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:HUMTOPII	+	170.00	131.79	41.82	4792	! J04088 Human DNA topoisomerase	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_ov:RNNHCB	+	170.00	129.91	53.23	5932	! Y15939 Rat mRNA for beta card	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_prio:AB025260	+	170.00	129.90	53.08	5932	! AB025260 Sus scrofa mRNA for m	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_hlg10:HUMBYH7CD	+	170.00	129.79	54.00	6008	! M58018 Homo sapiens beta-myosin	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_in2:AINHCHRC	+	170.00	128.76	61.74	6751	! X55714 Scallop (A. irradians) m	gb_hlg6:AC017670	+	169.50	101.81	2.0e+03	192290	! AC020761 Homo sapiens chr
gb_pat2:143748	+	170.00	125.85	89.60	93								



gb_hcg24:CEY52B11	-	178.00	103.66	1.5e+03	233941	298867	Caenorhabditis elegans	gb_in3:CELT04D1	-	175.00	119.89	192.52	28678	AF067617	Caenorhabditis el
gb_hcg1:AC006279	-	178.00	103.22	1.6e+03	245802	AC006279	Plasmodium falciparum	gb_hcg1:AC006710	-	175.00	105.48	1.2e+03	145920	AC006710	Caenorhabditis e
gb_hcg1:AC006702	-	178.00	101.46	2.0e+03	299864	AC006702	Caenorhabditis elegans	gb_pr2:AC006101	-	175.00	105.33	1.2e+03	148278	AC006101	cib338.f.24, c
gb_pr1:AB046824	+	177.50	141.58	11.92	3098	AB046824	Homo sapiens mRNA for	gb_in1:AC008218	-	175.00	103.68	1.5e+03	178792	AC008218	Drosophila melan
gb_in3:D63984	+	177.50	139.69	15.18	3832	D63984	Anthracidaris crassispin	gb_pr2:AC007461	-	175.00	103.60	1.6e+03	180385	AC007461	Homo sapiens chr
gb_pr9:HSN0AMR	+	177.50	134.16	30.86	7154	Z11583	H.sapiens mRNA for Numa	gb_in1:AC007892	+	175.00	102.55	1.8e+03	202929	AC007892	Drosophila melan
gb_pr9:HSN0DMR	+	177.50	129.26	57.89	12446	X14298	Human mRNA for dystrop	gb_in2:AE003770	-	175.00	101.67	2.0e+03	224100	AE003770	Drosophila melan
gb_hcg4:AC014744	-	177.50	122.25	142.21	27448	AC014744	Drosophila melanogast	gb_hcg24:CEY79H2	-	175.00	100.72	2.3e+03	249640	AL022287	Caenorhabditis e
gb_in1:AC005811	+	177.50	113.59	432.05	72987	AC005811	Drosophila melanogast	gb_hcg1:AC079433	-	175.00	100.61	2.3e+03	252689	AC079433	Mus musculus chr
gb_in2:AC003669	+	177.50	101.91	1.9e+03	272605	AC003669	Drosophila melanogast	gb_in3:CEY75B8A	-	175.00	99.14	2.8e+03	298406	AL033514	Caenorhabditis e
gb_pr7:BC006094	+	177.00	142.16	11.07	2775	BC006094	Homo sapiens, clone M	gb_hcg24:CEY75B8	-	175.00	98.07	3.2e+03	336638	AL022286	Caenorhabditis e
gb_pr1:AF026255	+	177.00	141.60	11.90	2957	AF026255	Mus musculus plenty-C	gb_in4:PFMEZSALB	-	174.50	134.43	29.81	5312	Z35327	P.falciparum gp190 (
gb_hcg2:G28541	+	177.00	138.40	17.93	4243	G28541	human STS SHGC-31621, S	gb_in4:PFMEZSALB	-	174.50	132.62	37.62	6519	AR016730	Sequence 1 from pa
gb_pr7:D63875	+	177.00	138.40	17.93	4243	G28541	human STS SHGC-31621, S	em_pat:EL11754	-	174.50	132.62	37.62	6519	EL11754	cDNA coding DOCK180
gb_pr1:AC003476	+	177.00	134.91	28.03	6287	AB003476	Homo sapiens mRNA for	gb_pr10:HUMDOCK180	-	174.50	132.62	37.62	6519	D50857	Human DOCK180 protei
gb_pr4:AF001504	+	177.00	134.06	28.08	6297	AF001504	Homo sapiens myasthen	gb_hcg1:AC079433	-	174.50	129.53	55.91	92883	AB035419	Drosophila melanog
gb_pr10:HUMBPA	+	177.00	134.06	31.26	6921	M63618	Human bullous pemphig	gb_v11:AF318573	+	174.50	107.68	922.10	103873	AF318573	Bovine herpesvir
gb_in2:AF221715	+	177.00	125.69	91.46	17802	AF221715	Drosophila melanogast	gb_hcg20:ALL59162	-	174.50	103.93	1.5e+03	156234	AL159162	Homo sapiens chr
gb_in2:AF184612	+	177.00	125.38	95.14	18430	AF184612	Drosophila melanogast	gb_hcg25:PFMAL13P3	-	174.50	98.17	3.1e+03	318221	AL049184	Plasmodium falci
gb_in2:AF188205	+	177.00	124.91	101.11	19444	AF188205	Drosophila melanogast	gb_pat1:AS8933	+	174.00	133.96	31.69	5361	AS8933	Sequence 2 from Pate
gb_hcg1:AC005646	-	177.00	111.67	552.52	86677	AC005646	Drosophila melanogast	gb_in4:PFAY010	-	174.00	133.68	32.82	5529	AJ007010	Plasmodium falcipa
gb_in1:AC005334	-	177.00	109.44	735.18	111448	AC005334	Drosophila melanogast	gb_pr4:SCYGR098C	-	174.00	133.32	34.40	5763	Z72883	S.cerevisiae chromos
gb_hcg9:AC024486	-	177.00	106.20	1.1e+03	160903	AC024486	Homo sapiens chromo	gb_in3:DMU77947	-	174.00	129.98	52.79	8401	U77947	Drosophila melanog
gb_in1:AC007580	-	177.00	106.19	1.1e+03	160903	AC007580	Drosophila melanogast	gb_hcg4:AC014378	-	174.00	125.33	95.86	14201	AC014378	Drosophila melanog
gb_in1:AC008371	-	177.00	106.05	1.1e+03	163461	AC008371	Drosophila melanogast	gb_hcg5:AC016449	-	174.00	125.33	95.86	14201	AC014378	Drosophila melanog
gb_hcg21:AC053535	+	177.00	105.87	1.2e+03	166826	AC053535	Homo sapiens chromo	gb_in1:AE003429	-	174.00	98.31	3.1e+03	299620	AE003429	Homo sapiens clon
gb_in2:AC008306	+	177.00	105.72	1.2e+03	166826	AC008306	Drosophila melanogast	gb_vv:AF111423	-	173.50	137.69	19.63	3364	AF111423	Xenopus laevis chr
gb_hcg8:AC023201	+	177.00	105.64	1.3e+03	171158	AC023201	Homo sapiens chromo	gb_pat1:AX023763	-	173.50	136.67	22.36	3773	AX023763	Sequence 20 from P
gb_hcg23:AL590413	+	177.00	105.10	1.3e+03	181935	AL590413	Homo sapiens chromo	gb_pr1:MMAP001794	-	173.50	136.13	23.99	4013	AF001794	Mus musculus Treac
gb_hcg6:AC019340	-	177.00	105.02	1.3e+03	183481	AC019340	Homo sapiens chromo	gb_r01:MMU81030	+	173.50	135.26	26.82	4428	U81030	Mus musculus treacle
gb_in1:AE003814	+	177.00	104.91	1.3e+03	260713	AE003814	Drosophila melanogast	gb_vv:FHV7G1	-	173.50	135.26	26.82	4428	U81030	Mus musculus treacle
gb_in1:AE003590	+	177.00	100.62	2.3e+03	301639	AE003590	Drosophila melanogast	gb_pr1:AB011125	-	173.50	133.84	32.18	5198	U07055	Fundulus heteroclitu
gb_in2:AF215635	+	176.50	132.42	38.59	7967	AF215635	Drosophila melanogast	gb_pr4:AF007217	-	173.50	133.22	34.84	5574	AF111125	Homo sapiens mRNA
gb_in2:AF214663	+	176.50	132.35	38.95	8033	AF214663	Drosophila melanogast	gb_pr1:AE002334	-	173.50	131.49	43.48	6411	AF007217	Homo sapiens Tri2
gb_in3:CELMAYNC	+	176.50	131.34	44.32	9000	J01050	Caenorhabditis elegans	gb_r01:MUSSPNA	+	173.50	130.56	48.99	7524	M74773	Mus musculus brain b
gb_hcg20:CEL158197	+	176.50	115.07	357.09	56457	AL158197	Homo sapiens chromos	gb_pr2:AC007722	-	173.50	110.11	675.21	75703	AC007722	Homo sapiens chro
gb_hcg24:CEL10666_3	+	176.50	109.16	761.95	110000	Continuation (4 of 7) of CEX		gb_hcg1:AC079414	-	173.50	102.89	1.7e+03	170916	AC079414	Homo sapiens chr
gb_pr2:ATF24M12	+	176.50	107.72	917.32	129516	AL132980	Arabidopsis thalian	gb_hcg6:AC019152	-	173.50	102.41	1.8e+03	180520	AC019152	Homo sapiens chr
gb_hcg20:AL353644	+	176.50	106.40	1.1e+03	129586	AL353644	Homo sapiens chromo	gb_hcg18:AC087226	-	173.50	98.97	2.8e+03	265985	AC087226	Mus musculus chr
gb_in1:AC010843	+	176.50	105.26	1.3e+03	170913	AC010843	Drosophila melanogast	gb_ba2:AC020295	-	173.00	96.62	3.8e+03	346897	AF002995	Mesohizobium lo
gb_hcg2:AC010994	+	176.50	104.53	1.4e+03	185573	AC010994	Drosophila melanogast	gb_r02:RNDNATPII	-	173.00	132.81	32.30	4987	Z19552	R.norvegicus mRNA fo
gb_hcg9:AC023477	+	176.50	104.02	1.5e+03	196536	AC023477	Homo sapiens chromo	gb_pr5:AF227948	+	173.00	133.23	34.78	5323	AF227948	Homo sapiens HBV P
gb_hcg13:AL355134	+	176.50	103.66	1.5e+03	204603	AC066591	Homo sapiens chromo	gb_r02:RNDNATPII	-	173.00	132.09	40.24	6052	Z46372	R.norvegicus RNA for
gb_hcg20:AL355134	+	176.50	103.12	2.1e+03	272677	AL355134	Homo sapiens chromo	gb_in3:CEY39847	-	173.00	117.62	257.55	31002	U39847	Caenorhabditis eleg
gb_in1:AE003492	+	176.50	100.31	2.4e+03	298839	AL353492	Drosophila melanogast	gb_hcg4:AC014947	-	173.00	116.56	295.10	34946	AC014947	Drosophila melanog
gb_pr10:HUMRGSLU	+	176.50	104.88	13.04	2933	M18740	Human chromosomal t(x;2	gb_in3:CELS0350	-	173.00	114.85	367.60	42400	U50071	Caenorhabditis eleg
gb_pr10:HUMRGSLU	+	176.00	136.93	21.64	4580	M25718	Human rDNA and 4 Alu re	gb_hcg6:AC020289	-	173.00	112.88	472.83	52916	AC020289	Drosophila melanog
gb_pr1:AR061132	+	176.00	136.67	22.37	4716	AR061132	Sequence 1 from pater	gb_in1:AC005977	-	173.00	110.49	642.73	69330	AC005977	Drosophila melanog
gb_r01:MWSOS1G	+	176.00	136.55	22.73	4783	Z11574	M.musculus mRNA for sor	gb_hcg2:AE001093	-	173.00	101.90	1.9e+03	182757	AC010993	Drosophila melan
gb_pr1:AB046853	+	176.00	136.06	24.20	5054	AB046853	Homo sapiens mRNA for	gb_in2:AE003816	-	173.00	98.75	2.9e+03	260849	AE003816	Drosophila melan
gb_pr5:AF111782	+	176.00	134.55	29.37	5992	AF111782	Homo sapiens extraocu	gb_in3:CTBRI	-	172.50	143.33	9.53	1629	X64322	C.tenacetans BRI gene.
gb_pr10:HSU13369	+	176.00	117.09	275.67	42999	U13369	Human ribosomal DNA cc	gb_r01:MUSSTOP2	+	172.50	133.96	31.66	4687	D12513	Mouse mRNA for DNA t
gb_pr6:AL157769	+	176.00	104.81	1.3e+03	172052	AL157769	Human DNA sequence	gb_r01:AF102129	-	172.50	132.77	36.92	5365	AF102129	Rattus norvegicus
gb_hcg24:AP003329	+	176.00	102.54	1.8e+03	222125	AP003329	Oryza sativa chromo	gb_pat1:EA04562	+	172.50	132.14	40.02	5760	A04562	P.falciparum P.195 g
gb_hcg24:CEY5485	-	176.00	100.65	2.3e+03	275138	Z29862	Caenorhabditis elegans	gb_pat2:EA00656	+	172.50	132.14	40.02	5760	E00556	cDNA encoding P.195
gb_hcg24:SFGAR2GN	+	175.50	143.73	9.05	2034	Z48166	S.pombe gar2 gene. 6/19	gb_vv:ASB039672	+	172.50	131.88	41.38	5932	ASB039672	Pennaria argentea
gb_pr5:AK026978	+	175.50	142.82	10.17	2254	AK026978	Homo sapiens cDNA: FL	gb_pr1:AF083424	-	172.50	130.54	49.10	6896	AB033071	Homo sapiens mRNA
gb_in2:AF145609	+	175.50	138.62	17.43	3620	AF145609	Drosophila melanogast	gb_v11:AF083424	-	172.50	106.13	1.1e+03	108409	AF083424	Ateline herpesvi
gb_r01:MUSP150R	+	175.50	137.30	20.64	4201	AL9502	Mus musculus phosphop	gb_hcg14:AC068881	-	172.50	105.78	1.2e+03	112833	AC068881	Homo sapiens chr
gb_r02:RUSP243304	+	175.50	137.22	20.86	4241	AJ243304	Rattus norvegicus mRN	gb_hcg22:AL390958	-	172.50	102.19	1.9e+03	156212	Z73551	S.cerevisiae chromos
gb_pr5:AF126008	+	175.50	137.11	21.13	4290	AF126008	Homo sapiens breast c	gb_pr4:SCYPL195W	-	172.00	136.38	23.22	3412	Z73551	S.cerevisiae chromos
gb_in4:SCABLAC	+	175.50	136.60	22.58	4548	E08855	Novel gene involved in	gb_pr3:SCABRI	-	172.00	136.24	23.63	3465	X51780	yeast ABPI gene for
gb_pr10:HSU13369	+	175.50	133.89	31.96	6173	L76606	Stenotocerca americana	gb_pr1:AB037861	-	172.00	134.06	31.27	4434	AB037861	Homo sapiens mRNA
gb_pr10:HUMMYOHC	+	175.50	132.05	40.45	7596	M69181	Human nonmuscle myosin	gb_pr5:HSBWHCF	+	172.00	134.05	31.30	4437	X51592	Human MHC mRNA for f
gb_r01:AF026032	+	175.50	131.96	40.92	7673	AF026032	Mus musculus ATRX pro	gb_pr5:AF125672	+	172.00	128.10	67.14	8686	AF125672	Homo sapiens sllen
gb_om:AF070485	+	175.50	126.71	80.29	13887	AF070485	Canis familiaris dys	gb_ba1:AE005123	-	172.00	125.85	89.64	11202	AE005123	Halobacterium sp.
gb_pr4:SPAC140	+	175.50	125.83	89.80	15326	AL163191	S.pombe chromosome I	gb_pr1:AE002920	-	172.00	121.89	148.99	17518	AE002920	Homo sapiens mRNA
gb_pr4:SPAC136	+	175.50	120.40	180.24	28294	Z70690	S.pombe chromosome I	gb_pr5:AF141968	-	172.00	121.78	151.08	17734	AF141968	Homo sapiens trab
gb_hcg1:AC005505	+	175.50	117.88	249.14	37621	Z68198	S.pombe chromosome I	gb_hcg6:AC012920	-	172.00	115.20	351.36	37274	AC012920	Drosophila melanog
gb_hcg1:AC005139	+	17													



gb_r01:AF013969	+ 185.50	143.73	9.05	4957	! AF013969 Mus musculus antigen	gb_ba3:MMU47023	+ 181.00	135.63	25.58	8285	! U47023 Methanococcus marip
gb_p10:HUMNP220	+ 185.50	141.23	12.47	6571	! D83032 Homo sapiens mRNA for n	gb_p14:YCESR1	+ 181.00	135.19	27.04	8700	! D11088 Saccharomyces cerevi
gb_in3:CELF332E10	+ 185.50	126.85	78.85	3318	! U41992 Caenorhabditis elegans	gb_sy:AF134159	- 181.00	131.85	41.52	12689	! AF134159 Expression shuttl
gb_pat1:AX063459	+ 185.00	149.17	4.50	2565	! AX063459 Sequence 4 from Paten	gb_sy:SC29711	+ 181.00	124.32	109.07	29886	! X75891 S.cerevisiae 29.7Kb
gb_in2:AF042160	+ 185.00	147.79	5.38	2598	! AF042160 Plasmodium falciparum	gb_in3:CELUK03C7	+ 181.00	124.23	110.34	29989	! U40059 Caenorhabditis eleg
gb_in2:AB042827	+ 185.00	147.37	5.67	3142	! AB042827 Rattus norvegicus mRN	gb_htg18:AC087689	+ 181.00	106.60	1.1e+03	219259	! X15309 H.sapiens NF-H gene,
gb_ov:PMU19361	+ 185.00	145.78	6.96	3763	! U19361 Petromyscus marinus neur	gb_pr3:HSNFH4	+ 180.50	143.97	7.03	2542	! L06487 Saccharomyces cerevi
gb_r01:AF051357	+ 185.00	144.40	8.30	4393	! AF051357 Mus musculus golgin-2	gb_p14:YSCZ1P1A	+ 180.50	145.77	8.78	3091	! M69183 Plasmodium falciparu
gb_om:SSU75316	+ 185.00	141.65	11.82	5996	! U75316 Sus scrofa beta-myosin	gb_in4:PFAMESA	+ 180.50	139.71	15.14	4995	! L40157 Human endosome-associ
gb_ba1:AE006525	+ 185.00	137.09	21.20	10029	! AE006525 Streptococcus pyogen	gb_pr10:HUMERAP	+ 180.50	139.66	15.26	5028	! AF285157 Gallus gallus topo
gb_p10:HSU91931	+ 184.50	143.84	6.90	3574	! U91931 Homo sapiens AP-3 compl	gb_ov:AF285155	+ 180.50	139.43	15.71	5158	! AB015484 Dugesia japonica m
gb_pr3:HSP162	+ 184.50	143.26	9.61	4780	! X78998 H.sapiens p162 mRNA. 3/	gb_in1:AB015484	+ 180.50	138.05	18.75	6029	! AF166170 Caenorhabditis ele
gb_ov:AB007445	+ 184.50	141.58	11.92	5781	! AB007445 Gallus gallus mRNA f	gb_in2:AF166170	+ 180.50	135.94	24.56	7646	! AF166169 Caenorhabditis ele
gb_r01:MUSDY5A	+ 184.50	133.86	32.09	13815	! M68859 Mouse dystrophin mRNA,	gb_in2:AF166169	+ 180.50	135.79	25.06	7781	! AF261891 Caenorhabditis ele
gb_pat1:AR093392	+ 184.50	130.89	46.94	19307	! AC093392 Sequence 10 from Pat	gb_in2:AR261891	+ 180.50	134.96	27.86	8543	! AF160252 Homo sapiens KIAA
gb_p11:AC005617	+ 184.50	125.42	94.74	35822	! AC005617 Arabidopsis thaliana	gb_pr3:AF160252	+ 180.50	132.24	39.49	11611	! U51031 Saccharomyces cerev
gb_p10:NCB13118	+ 184.50	121.01	166.76	58918	! AL390189 Neurospora crassa DN	gb_htg6:AC020215	+ 180.50	125.47	94.14	24942	! AC020215 Drosophila melano
gb_htg16:AC079219	+ 184.50	111.07	596.97	181008	! AC079219 Mus musculus chromo	gb_pr3:AC005527	+ 180.50	110.48	238.43	56510	! AC000025 Homo sapiens Chr
gb_htg6:AC019525	+ 184.50	109.72	709.29	210663	! AC019525 Drosophila melanoga	gb_p11:AC025806	+ 180.50	109.26	752.70	155420	! AC025806 Oryza sativa chr
gb_pr6:AL3355497	+ 184.50	109.52	727.43	215397	! AL335497 Human DNA sequenc	gb_in1:AE003351	+ 180.50	103.86	1.1e+03	285978	! AE003351 Drosophila melan
gb_in1:AE0033481	+ 184.50	106.02	1.1e+03	319850	! AE0033481 Drosophila melanoga	gb_in1:AE003351	+ 180.50	103.51	1.6e+03	297385	! U50351 Saccharomyces cerev
gb_p14:SCYL1021W	+ 184.00	141.86	11.50	5354	! X73126 S.cerevisiae chromosome	gb_pr3:SC005529	+ 180.50	102.90	1.7e+03	134308	! X59720 S.cerevisiae chrom
gb_p13:SCSPAG	+ 184.00	140.79	13.19	6041	! X53731 S.cerevisiae SPA2 gene	gb_pat2:ELJ401	+ 180.00	146.37	6.45	2254	! E14401 cDNA encoding HET-B
gb_ov:AR072938	+ 184.00	136.14	23.96	10216	! U72938 Homo sapiens putative	gb_p10:HSU81504	+ 180.00	141.40	12.20	3950	! U81504 Homo sapiens beta-3A
gb_p10:HSU72937	+ 184.00	136.04	24.26	10330	! U72937 Homo sapiens putative	gb_in2:AR247194	+ 180.00	139.70	15.17	4784	! AF247194 Drosophila melanog
gb_r01:HSU72936	+ 184.00	135.94	24.59	10452	! U72936 Homo sapiens putative	gb_om:OCU32574	+ 180.00	137.71	19.57	5988	! U32574 Oryctolagus cuniculu
gb_in2:AR297644	+ 184.00	135.48	26.06	11000	! AF297644 Drosophila melanogas	gb_pr3:HSU53786	+ 180.00	137.04	21.32	6457	! U53786 Homo sapiens envlopi
gb_p13:SCG32KBP	+ 184.00	126.00	87.94	32088	! X97560 S.cerevisiae 32kb DNA	gb_in3:HSU91980	+ 180.00	136.06	24.20	7231	! U11584 H.sapiens mRNA for N
gb_ov:GGNFM1D	+ 183.50	141.01	12.83	5638	! X17102 Chicken DNA for middle-	gb_pat1:AR031367	+ 180.00	134.77	28.56	8351	! AE031367 Sequence 14 from p
gb_in1:AE001424	+ 183.50	134.14	30.95	12240	! AE001424 Plasmodium falciparu	gb_pat2:168238	+ 180.00	134.77	28.56	8351	! 168238 Sequence 14 from pat
gb_p13:NCB1109	+ 183.50	118.56	228.40	71085	! AL513409 Neurospora crassa DN	gb_p11:HSU75653	+ 180.00	132.86	36.48	10358	! U75653 Human zinc finger h
gb_htg14:AC068279	+ 183.50	112.77	479.57	136555	! AC068279 Homo sapiens chromo	gb_v11:AF162221	+ 180.00	107.63	928.05	178733	! AF162221 Xestia c-nigrum
gb_om:AR0234765	+ 183.00	152.43	2.96	1485	! AF234765 Rattus norvegicus sex	gb_st52:ONS07DYV	+ 179.50	153.25	2.67	992	! AL441104 T7 end of clone B00
gb_om:AR0234765	+ 183.00	140.16	14.31	5938	! AB234765 Sus scrofa MYHC-s1	gb_pr3:BC004286	+ 179.50	140.50	13.68	4179	! BC004286 Homo sapiens, clon
gb_pat2:AX068804	+ 183.00	139.65	15.26	6284	! AX068804 Sequence 756 from Pat	gb_pr5:AF214114	+ 179.50	137.59	19.87	5804	! AF214114 Homo sapiens, retin
gb_pr3:HS0801013	+ 183.00	139.65	15.26	6284	! AL117496 Homo sapiens mRNA; ch	gb_om:AB059400	+ 179.50	137.40	20.36	5930	! AB059400 Bos taurus MYHC-s1
gb_in3:CEH38K22	+ 183.00	134.05	31.31	11829	! AE001376 Plasmodium falciparu	gb_pat1:AR018129	+ 179.50	137.20	20.91	6069	! U42391 Human myosin-IX mRN
gb_htg3:AC073693	+ 183.00	127.21	75.32	25610	! AL024499 Caenorhabditis elega	gb_pat2:173445	+ 179.50	136.86	21.84	6306	! AF018129 Sequence 3 from pa
gb_p10:HSU91930	+ 182.50	107.74	913.93	230372	! AC073693 Mus musculus clone	gb_pat2:185731	+ 179.50	136.86	21.84	6306	! I85731 Sequence 3 from pate
gb_p13:SCU17580	+ 182.50	141.75	11.66	4744	! U91930 Homo sapiens AP-3 compl	gb_pr3:AC000035	+ 179.50	120.85	170.24	38429	! AC000035 Homo sapiens Chro
gb_p10:HUMYTOHP	+ 182.50	139.85	14.88	5877	! U19580 Saccharomyces cerevisia	gb_htg16:AC074305	+ 179.50	105.77	1.2e+03	216677	! AC074305 Mus musculus chr
gb_pat1:AR019266	+ 182.50	139.64	15.28	6016	! M36769 Homo sapiens perinatal	gb_htg20:AL355134	+ 179.50	103.49	1.6e+03	272677	! AL355134 Homo sapiens chr
gb_htg4:AC012625	+ 182.50	139.23	16.12	6306	! AC012625 Homo sapiens chromo	gb_in1:AE003543	+ 179.50	103.12	1.7e+03	284398	! AE003543 Drosophila melan
gb_r01:AC018634	+ 182.50	111.93	534.24	137363	! AC018634 Human Chromosome 7	gb_r01:MMU63648	+ 179.00	140.22	14.19	4126	! U63648 Mus musculus p160 my
gb_r01:AC018634	+ 182.50	109.82	699.99	174241	! D63850 Mus musculus mRNA for h	gb_pr3:AB019691	+ 179.00	130.56	48.98	12278	! AB019691 Homo sapiens mRNA
gb_in2:AF145640	+ 182.00	147.92	5.28	2260	! AF145640 Drosophila melanogast	gb_p13:SC29582X	+ 179.00	120.40	180.30	38661	! Z49259 S.cerevisiae chromo
gb_r01:AF340028	+ 182.00	142.37	10.77	4230	! AF340028 Mus musculus Rab6-int	gb_p13:NC99H12	+ 179.00	113.40	442.31	85165	! AL451018 Neurospora crassa
gb_pr3:AB015617	+ 182.00	140.99	12.86	4943	! AB015617 Homo sapiens ELKS mRN	gb_htg1:AC009091	+ 179.00	106.24	1.1e+03	191140	! AC009091 Homo sapiens chr
gb_in2:AF134172	+ 182.00	140.36	13.94	5307	! Y11251 H.sapiens mRNA for nove	gb_r02:RNAJ2367	+ 178.50	131.38	44.09	10705	! AC002967 Rattus norvegicus
gb_ov:AF004397	+ 182.00	139.48	15.61	5863	! AF134172 Pecten maximus myosin	gb_htg4:AC014439	+ 178.50	125.56	93.04	20656	! AC014439 Drosophila melano
gb_r01:AC000015	+ 182.00	138.07	18.69	6872	! AF004397 Gallus gallus chromo-	gb_p13:ART28119	+ 178.50	110.68	627.21	110766	! AC035709 Arabidopsis thal
gb_htg24:AP003235	+ 182.00	117.94	247.32	66704	! AC000015 Homo sapiens chromo	gb_htg12:AC044873	+ 178.50	106.01	1.1e+03	187554	! AC044873 Homo sapiens chr
gb_htg18:AC084878	+ 182.00	110.47	644.54	154975	! AP003235 Oryza sativa chromo	gb_htg1:AC079501	+ 178.50	105.51	1.4e+03	198493	! AL61592 Arabidopsis thal
gb_in1:AC008216	+ 182.00	109.81	701.02	179486	! AC008478 Homo sapiens chromo	gb_htg1:AC079501	+ 178.50	104.65	1.4e+03	218687	! AC079501 Mus musculus cio
gb_htg2:AP003286	+ 182.00	109.17	761.57	169486	! AC008216 Drosophila melanoga	gb_in3:SDNF220A	+ 178.00	139.55	15.47	4074	! M94389 Squid neurofilament
gb_in2:AE0033759	+ 182.00	109.08	769.84	181200	! AP003286 Oryza sativa chromo	gb_pat1:AB018346	+ 178.00	136.67	22.39	5639	! AB018346 Homo sapiens mRNA
gb_htg2:AC010172	+ 182.00	107.11	991.86	226471	! AC003759 Drosophila melanoga	gb_in2:AS58932	+ 178.00	135.89	24.71	6152	! AS58932 Sequence 1 from Pate
gb_pr3:AF048977	+ 181.50	143.17	9.72	3698	! AF048977 Homo sapiens Ser/Arg-	gb_in2:AR324831	+ 178.00	132.44	38.52	9091	! AF324831 Plasmodium falcipa
gb_r01:AF043102	+ 181.50	142.47	10.64	4002	! AF043102 Pneumocystis carinii	gb_in3:HSANK8440	+ 178.00	129.60	55.40	12518	! Z26634 Homo sapiens mRNA f
gb_ba1:AF143911	+ 181.50	139.28	16.02	5737	! AF143911 Mycoplasma pneumoniae	gb_in3:CELUK09G12	+ 178.00	121.29	160.82	331979	! AF047663 Caenorhabditis el
gb_pr3:HSWYOSIN	+ 181.50	138.86	16.89	6010	! Z38133 H.sapiens mRNA for myos	gb_in3:CFE2809	+ 178.00	120.83	170.72	33706	! Z81518 Caenorhabditis eleg
gb_ba1:SCWYOIG	+ 181.50	138.72	17.20	6108	! X53947 Yeast MYO1 gene for myo	gb_in3:CELC42D8	+ 178.00	118.71	223.96	420001	! U56966 Caenorhabditis eleg
gb_ba1:AF143912	+ 181.50	137.44	20.28	7080	! AF143912 Mycoplasma pneumoniae	gb_htg24:CEV53HL1	+ 178.00	110.35	654.59	118001	! Continuation (2 of 4) of
gb_p10:HUMDYS	+ 181.50	131.40	43.99	13957	! M18533 Homo sapiens dystrophin	gb_htg1:AC013398	+ 178.00	106.58	1.1e+03	168216	! AC013398 Homo sapiens cio
gb_in3:CELY5AE10A	+ 181.50	113.81	419.83	101640	! AC024810 Caenorhabditis eleg	gb_htg11:AC026495	+ 178.00	105.67	1.2e+03	186506	! AC026495 Homo sapiens chr
gb_htg9:AC023572	+ 181.50	111.19	587.58	136630	! AC023572 Homo sapiens chromo	gb_htg11:AC010321	+ 178.00	104.53	1.4e+03	212074	! AC010321 Homo sapiens chr
gb_p11:AR002530	+ 181.00	142.72	10.30	3720	! AB002530 Neurospora crassa mus						
gb_pr4:AF091711	+ 181.00	139.00	16.59	5660	! AF091711 Homo sapiens splice v						
gb_pat1:AR002603	+ 181.00	137.63	19.77	6605	! AR002603 Sequence 4 from Paten						
gb_p10:HSU81607	+ 181.00	136.23	23.68	7740	! U81607 Homo sapiens gravin mRN						
gb_p14:SCYBRL136W	+ 181.00	136.23	23.68	7740	! Z36005 S.cerevisiae chromosome						



gb.p13:SCNUM1	202.00	150.67	3.72	9851	! X61236 S.cerevisiae NUM1 gene,	gb.ov:AF027728	192.00	142.99	9.94	9610	! AF027728 Xenopus laevis kin
gb.p13:SC8358	201.00	137.52	20.07	3468	! U30046 S.cerevisiae chromosom	gb.hg12:CBRG42F08	192.00	129.32	49.99	4999	! AC084596 Caenorhabditis br
gb.p13:SC8358	201.50	153.35	2.64	6965	! U51906 Homo sapiens golgin-245	gb.hg18:AC087748	192.00	125.44	94.52	69740	! AC087748 Homo sapiens chr
gb.pat2:135495	201.50	140.82	4.04	10136	! I35495 Sequence 2 from patent	gb.hg4:AC012857	192.00	117.64	266.79	168062	! AC012857 Drosophila melan
gb.p10:1HUMGCP372	201.50	149.08	4.11	10295	! D25542 Human mRNA for goli	gb.pr3:AC010723	192.00	117.31	267.85	174420	! AC010723 Homo sapiens BAC
gb.p14:SPRC27	201.00	148.95	4.63	10943	! AL353866 S.pombe chromosome	gb.ro1:AF194970	191.50	143.81	8.96	8385	! AF194970 Mus musculus BAC1
gb.in4:PARHPR	200.50	152.52	2.93	6991	! L27838 Plasmodium yoelii rhop	gb.in1:AE003558	191.50	112.93	470.22	273744	! AE003558 Drosophila melan
gb.in2:AF335500	200.00	152.80	2.83	6484	! AF335500 Physarum polycephalum	gb.in1:AF0340183	191.00	149.02	4.59	4455	! AF340183 Homo sapiens BAC
gb.in2:AF323443	200.00	150.67	3.71	8241	! AF323443 Plasmodium yoelii yo	gb.pr9:HSMP63	191.00	141.59	11.91	10302	! X69086 H.sapiens mRNA for
gb.p14:SPCC162	200.00	137.14	21.07	37960	! AF323860 S.pombe chromosome	gb.pat1:AF343607	191.00	141.57	11.93	10320	! AF3607 Sequence 9 from Pat
gb.p11:AB0016871	200.00	130.63	48.53	79109	! BC001687 Arabidopsis thaliana	gb.v12:U97553	191.00	119.88	192.78	119450	! U97553 Murine herpesvirus
gb.pr7:BC001883	199.50	160.97	0.9911	2464	! BC001883 Homo sapiens, nucleol	gb.v11:AF105037	191.00	119.87	192.96	119550	! AF105037 Murid herpesviru
gb.pr7:BC003993	199.50	158.97	1.28	3089	! BC003993 Mus musculus, clone M	gb.hg12:AC037485	191.00	117.66	256.37	135317	! AC037485 Homo sapiens chr
gb.in2:AF196347	199.50	156.32	1.80	4164	! AF196347 Plasmodium falciparum	gb.in1:AC007853	191.00	117.13	274.29	182921	! AC007853 Drosophila melan
gb.in2:AF320345	199.50	118.35	234.65	302855	! AC003420 Drosophila melanoga	gb.in1:AC008206	191.00	116.19	309.38	181132	! AC008206 Drosophila melan
gb.ov:AF286475	199.00	160.90	1.00	2376	! AF286475 Takifugu rubripes ret	gb.in2:AE003750	191.00	114.17	400.64	227398	! AE003750 Drosophila melan
gb.in2:AF134186	199.00	155.59	1.98	4326	! AF134186 Caenorhabditis elegans	gb.ba3:NNAG622491	191.00	110.98	603.88	326301	! AL162757 Neisseria mening
gb.ro2:CATGCP60	199.00	148.02	5.22	10170	! D25543 Rattus norvegicus mRNA	gb.om:AF165917	190.50	150.39	3.85	3649	! AF165917 Canis familiaris t
gb.in3:CELT03A1	199.00	136.99	21.48	35318	! AF067614 Caenorhabditis eleg	gb.in3:DR00XONNA	190.50	145.50	7.21	6338	! M35012 D.melanogaster gon-m
gb.in1:AC008140	199.00	124.73	103.53	140973	! AC008140 Drosophila melanoga	gb.pr10:HUMSEQX	190.50	141.99	11.31	9416	! L06237 Human microtubule-as
gb.in1:AC009219	199.00	123.66	118.70	159007	! AC009219 Drosophila melanoga	gb.hg6:AC020480	190.50	138.81	17.01	13487	! AC020480 Drosophila melan
gb.in3:CEU010412	198.50	154.66	2.23	4597	! U10412 Caenorhabditis elegans	gb.in2:CEC25A1	190.50	129.04	59.49	40597	! Z81038 Caenorhabditis eleg
gb.in1:AC008140	198.00	152.46	2.95	5894	! Y00624 Acanthamoeba castellan	gb.in3:DME271740	190.50	124.17	111.20	70398	! AJ271740 Drosophila melan
gb.in2:AF134186	198.00	152.13	3.08	5850	! U62026 Gallus gallus cardiac m	gb.hg6:AC017682	190.50	123.57	120.07	75313	! AC017682 Drosophila melan
gb.ov:GGU62026	197.50	146.72	6.16	10300	! X75304 H.sapiens giantin mRNA	gb.hg24:CEX3986_1	190.50	120.21	184.66	110000	! Continuation (2 of 4) of
gb.pat2:181218	197.00	147.73	5.41	8789	! I81218 Sequence 5 from patent	gb.p11:AC009323	190.50	117.97	256.32	141748	! AC009323 Arabidopsis thal
gb.pr10:HSU030872	197.00	146.40	6.42	10211	! U30872 Human mitotin mRNA, co	gb.hg4:AC012609	190.50	117.56	259.45	148380	! AC012609 Homo sapiens chr
gb.hg18:AC080675	197.00	119.97	190.43	184494	! AC087189 Homo sapiens chromo	gb.ov:GCLCAUSA	189.50	150.30	3.89	3371	! X048278 G.gallus mRNA for cl
gb.in3:CEU012F3	196.50	165.74	0.5375	36555	! U80022 Caenorhabditis elegans	gb.pat1:AR048216	189.50	147.05	5.91	4868	! AR048216 Sequence 12 from p
gb.om:OCU34201	196.00	130.31	1.08	2033	! U34201 Oryctolagus cuniculus	gb.hg7:AC020738	189.50	116.65	291.75	164520	! AC020738 Homo sapiens clo
gb.ba1:AF053765	196.00	147.62	5.49	8142	! AF053765 Bacillus megaterium	gb.hg9:AC024601	190.50	116.26	306.58	171857	! AC024601 Homo sapiens chr
gb.pr10:HSU019769	196.00	145.71	7.01	10096	! U19769 Human CENP-F kinetoch	gb.in3:CEY3986B	190.50	114.61	378.80	207022	! AL132896 Caenorhabditis e
gb.in3:DMC49E4	196.00	133.57	33.28	37949	! AL031128 Drosophila melanogas	gb.pr2:AC008174	190.00	120.35	181.58	103663	! AC008174 Homo sapiens BAC
gb.hg18:AC080675	196.00	119.97	190.43	184494	! AC087189 Homo sapiens chromo	gb.pr10:HSU18985	189.50	151.49	3.34	2947	! U18985 Human triadin mRNA,
gb.sts1:AC080675	195.50	165.74	0.5375	36555	! U80022 Caenorhabditis elegans	gb.ov:GCLCAUSA	189.50	150.30	3.89	3371	! X048278 G.gallus mRNA for cl
gb.ba1:AB002503	195.50	144.20	8.52	11455	! AC02503 Neisseria meningitid	gb.pat1:AR048216	189.50	147.05	5.91	4868	! AR048216 Sequence 12 from p
gb.in3:CEU012F3	195.50	136.53	22.79	72453	! Z78066 Caenorhabditis elegans	gb.in3:DME18278	189.50	133.97	14.65	7079	! AF003622 Drosophila melanog
gb.p12:AF0203057	195.50	127.75	70.28	73391	! AP020305 Arabidopsis thaliana	gb.hg2:AC009160	189.50	114.54	382.17	130856	! AC009160 Homo sapiens chr
gb.p11:AB018113	195.50	126.44	83.06	85020	! Z38059 S.cerevisiae chromosom	gb.ba2:ECR75137A	189.00	146.53	6.32	4936	! X59794 E.coli broad-host-ra
gb.pat1:AX044033	195.00	113.91	414.63	349980	! AX044033 Sequence 112 from F	gb.pr4:AF083037	189.00	139.18	16.22	11318	! AF233527 Arabidopsis thalia
gb.om:BTCYCN1	195.00	159.24	1.24	2007	! Z46789 B.taurus mRNA for cylv	gb.pr8:HSAL31693	189.00	138.32	18.10	12465	! AJ131693 Homo sapiens mRNA
gb.hg1:AC006280	195.00	154.86	2.17	3289	! AK026473 Homo sapiens cDNA: FL	gb.p14:SPAC1F5	189.00	127.94	68.56	40242	! Z68136 S. pombe chromosome
gb.pr5:AK026473	195.00	120.26	183.62	163443	! AC06280 Plasmodium falcipar	gb.ba2:BAU67194	189.00	125.44	94.44	53339	! U67194 Enterobacter aeroge
gb.hg1:AC006281	195.00	118.49	230.36	195551	! AC06281 Plasmodium falcipar	gb.hg6:AC019755	189.00	122.01	146.70	78598	! AC019755 Drosophila melan
gb.om:RADNAA	194.50	157.22	1.60	2410	! M94315 Rabbit neurofilament-H	gb.hg2:AC010344	189.00	116.12	312.31	152829	! AC010344 Homo sapiens chr
gb.om:RADNAA	194.50	151.00	3.56	4655	! L10065 Oryctolagus cuniculus	gb.p12:AF233527	188.50	148.18	5.11	3916	! AF233527 Arabidopsis thalia
gb.pr3:AFU07030	194.00	150.18	3.96	5103	! U47030 Rhizobium tropici putat	gb.om:AF176816	188.50	145.79	6.95	5134	! AF176816 Bos taurus KIA037
gb.om:AB009287	194.00	149.16	4.51	5724	! AB009387 Sus scrofa mRNA for	gb.hg9:AC024036	188.50	114.20	399.37	181485	! AC024036 Homo sapiens chr
gb.p1:AB002371	194.00	148.80	4.72	5967	! AB002371 Human mRNA for KIAA03	gb.ro2:RNU83590	188.00	147.98	5.25	3834	! U83590 Rattus norvegicus PA
gb.hg12:AC036204	194.00	120.12	186.92	151872	! AC036204 Homo sapiens chromo	gb.om:AB025761	188.00	144.12	8.61	5929	! AB025761 Sus scrofa mRNA fo
gb.pr3:AC009487	194.00	119.56	200.78	181742	! AC036204 Homo sapiens chromo	gb.in2:AF233442	188.00	141.18	12.54	8256	! AF233442 Plasmodium yoelii
gb.in2:AE003659	194.00	115.33	345.67	260897	! AC009487 Homo sapiens BAC cl	gb.ro1:MMY12229	188.00	138.56	17.55	11096	! Y12229 M.musculus mRNA for
gb.in2:AF217802	193.50	130.92	3.60	4490	! AF217802 Drosophila melanoga	gb.p13:F23A5	188.00	118.27	237.09	109694	! AC011713 Arabidopsis thal
gb.v12:HSV3PRGEN	193.50	150.77	47.69	43658	! M86409 Herpesvirus salmire th	gb.pr9:HSGMAP210	187.50	155.70	1.95	1533	! M94315 Rabbit neurofilament
gb.hg4:AC015193	193.50	130.54	49.13	44815	! AC015193 Drosophila melanogas	gb.pr9:BSGMAP210	187.50	142.97	9.97	6452	! Y12490 Homo sapiens mRNA fo
gb.hg9:AC013932	193.50	122.31	141.19	112930	! X64346 Herpesvirus salmire c	gb.in3:DDIMVHC	187.50	142.66	10.37	6881	! M14628 D.discoideum myogin
gb.hg1:AC008284	193.50	120.03	189.17	1246797	! AC008284 Drosophila melanoga	gb.ba1:MPU59896	187.50	137.41	20.34	12083	! U59896 Mycoplasma pneumonia
gb.in2:AE003752	193.50	116.22	308.07	225748	! AC003752 Drosophila melanoga	gb.ro1:WUSNLF5A5	187.50	139.30	15.96	8930	! M69225 Human bullous pemph
gb.sts1:CN50797T	193.00	164.33	0.6441	945	! AL434943 T7 end of clone BB0A0	gb.in1:AE003091	186.50	130.57	48.90	23919	! AC003091 Drosophila melan
gb.sy:CVBL131294	193.00	149.68	4.22	4940	! A92451 Sequence 2 from Patent	gb.hg5:AC016135	186.50	115.53	336.64	130659	! AC009368 Homo sapiens chr
gb.pr4:AFU030234	193.00	149.06	4.57	5298	! AFU030234 Homo sapiens splicing	gb.pr7:BC000755	186.50	150.91	3.60	2305	! BC000755 Homo sapiens chr
gb.ba3:MHMP343G	193.00	146.85	6.06	6797	! X95601 M.hominis lmp3 and lmp4	gb.in2:AF245512	186.00	148.08	8.65	4380	! AF245512 Rattus norvegicus
gb.hg18:AF184623	193.00	143.36	9.48	10077	! AF184623 Plasmodium vivax ret	gb.in4:FFARBP4CA	186.00	142.60	10.46	5888	! FFARBP4CA Hydra vulgaris tig
gb.in3:CELT24973	192.50	130.30	50.67	42126	! AC090081 Homo sapiens chromo	gb.in1:AF087076	186.00	119.10	21.90	83495	! AC087076 Caenorhabditis br
gb.in1:AC084463	192.00	124.01	113.48	83644	! AC084463 Caenorhabditis eleg	gb.ro1:MUSNPH	185.50	145.72	7.01	3360	! M35131 Mouse neurofilament
gb.ov:XLNOPP180	192.00	152.12	3.09	3432	! X88927 X.laevis mRNA for XNOP	gb.ro1:MUSNPH	185.50	145.72	7.01	3360	! M35131 Mouse neurofilament
gb.in1:AB004305	192.00	147.27	4.44	4730	! AB004305 Mus musculus mRNA for	gb.ro1:MUSNPH	185.50	145.72	7.01	3360	! M35131 Mouse neurofilament
gb.ov:CCU53862	192.00	147.27	5.77	5957	! U53862 Coturnix coturnix slow	gb.ro1:MUSNPH	185.50	145.72	7.01	3360	! M35131 Mouse neurofilament
gb.in1:AF139185	192.00	145.83	6.91	6975	! AF139185 Rattus norvegicus myd	gb.ro1:MUSNPH	185.50	145.72	7.01	3360	! M35131 Mouse neurofilament



Sequence	Strnd	Orig	zScore	EScore	Len	Documentation
gb.pr8.HS26H231	+	7155.00	5641.68	5.2e-306	74444	! AL137201 Novel human gene mapped
gb.pr8.HS095825	+	7101.00	5602.15	8.3e-304	5253	! U95825 Human androgen-induced
gb.pr1.AB023196	+	6994.00	5517.93	4.1e-299	5110	! AB023196 Homo sapiens MRNA for
gb.pr10.HS050533	+	2953.50	2337.44	5.8e-132	1852	! U50533 Human BRCA2 region, mRNA
gb.pr1.AB014548	+	7274.50	2331.78	1.7e-110	5177	! AB014548 Homo sapiens MRNA for
gb.in1.AC007475	+	1729.50	2331.02	6.7e-66	173613	! AC007475 Drosophila melanogaster
gb.in1.AC007474	+	1729.50	1330.09	7.5e-66	192763	! AC007474 Drosophila melanogaster
gb.in2.AC003823	+	1729.50	1327.35	1.1e-65	262731	! AC003823 Drosophila melanogaster
gb.htg2.AC002086	+	1708.50	1313.43	6.4e-65	194634	! AC002086 Drosophila melanogaster
gb.pr8.AK026989	+	925.50	735.56	9.8e-33	2079	! AK026989 Homo sapiens CDNA: FL
gb.pr8.HSA49J10	+	925.50	698.45	1.1e-30	137246	! 284572 Human DNA sequence fr
gb.pr5.AL138820	+	925.50	696.63	1.4e-30	168487	! AL138820 Human DNA sequence
gb.pr14.SPAC	+	833.50	649.04	6.5e-28	99382	! AL441624 S. pombe chromosome I
gb.htg14.AC068224	+	819.50	622.97	1.8e-26	54398	! AC068224 Homo sapiens chromosome
gb.htg32.AL512630	+	769.00	566.85	2.4e-23	340969	! AL512630 Mus musculus chromosome
gb.htg5.AC016449	+	768.00	579.81	4.6e-24	72157	! AC016449 Homo sapiens clone R
gb.pr5.AK021757	+	696.50	562.14	4.5e-23	2212	! AK021757 Homo sapiens cDNA FL
gb.pr4.SMAJ9934	+	697.50	545.61	3.7e-22	6411	! AJ009934 Sordaria macrospora S
gb.pr13.EMB1ND	+	635.00	498.68	1.5e-19	4895	! L03200 Emericella nidulans bld
gb.htg14.AC068224	+	554.50	413.79	8.2e-15	54398	! AC068224 Homo sapiens chromosome
gb.htg14.AC068352	+	466.00	335.21	2.0e-10	145450	! AC068352 Homo sapiens chromosome
gb.htg21.A1358892	+	401.50	281.68	1.9e-07	195380	! A1358892 Mus musculus chromosome
gb.pr8.HS267P19	+	390.50	277.80	3.1e-07	113704	! 275889 Human DNA sequence fr
gb.htg30.AL153724	+	390.50	274.57	4.7e-07	163642	! AL153724 Homo sapiens chromosome
gb.pr13.FSC3582X	-	377.00	276.70	3.5e-07	38661	! 249259 S. cerevisiae chromosome
gb.pr13.AC023592	+	375.50	304.37	1.0e-08	1489	! AK023592 Homo sapiens cDNA FL
gb.htg8.AC023006	+	353.50	245.41	2.0e-05	162775	! AC023006 Homo sapiens clone
gb.htg14.AC068352	+	331.00	228.65	0.0002	145450	! AC068352 Homo sapiens chromosome
gb.htg18.AC090081	+	326.00	223.44	0.0003	167671	! AC090081 Homo sapiens chromosome
gb.htg8.AC022862	+	314.00	214.33	0.0011	160962	! AC022862 Homo sapiens chromosome
gb.pl1.AF049529	+	309.50	252.62	7.8e-06	1431	! AF049529 Schizosaccharomyces f
gb.htg40.AL135724	+	277.00	184.98	0.0456	763642	! AL135724 Homo sapiens chromosome
gb.pr10.AB019602	+	248.50	190.02	0.0239	13608	! AB019602 Homo sapiens IDN3-B m
gb.pr1.AB019494	+	232.50	184.68	0.0474	7821	! AB019494 Homo sapiens IDN3 mRN
gb.pr4.YSCMLP	+	246.50	180.96	0.0764	6978	! L01992 Saccharomyces cerevisiae
gb.pr13.SCDNRO95W	+	235.50	178.90	0.0994	8050	! Z78320 S. cerevisiae chromosome
gb.pr13.SCDNRO28X	+	235.50	172.92	0.2142	15820	! X73541 S. cerevisiae DNA of ch
gb.in3.CEBU0321	+	234.00	163.75	0.6940	38946	! U23168 Caenorhabditis elegans
gb.in2.AF1312916	+	232.50	176.59	0.1684	9800	! AF1312916 Plasmodium falciparum
gb.ov.XLNU0APRO	+	231.50	174.79	0.1338	7321	! Y07624 X.laevi mRNA for NuMA



**THIS PAGE BLANK (USPTO)**







shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 267P19 is at 1 in this sequence. The true right end of clone 26H23 is at 24539.

The true left end of clone 49J10 is at 113606.

267P19 is from the human PAC library.

#### FEATURES

##### Source

1. .113704

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="13"

/map="13q12-13"

/clone="XX-267P19"

1. .235

repeat\_region

repeat\_region /note="Alu repeat: matches 252. .1 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

repeat\_region /note="2 copies of 54 mer 87 % conserved"

repeat\_region

repeat\_region /note="MER7A element fragment"

repeat\_region

repeat\_region /note="L1 element fragment"

repeat\_region

repeat\_region /note="L1 element fragment"

repeat\_region

repeat\_region /note="Alu repeat: matches 305. .210 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 158. .1 of consensus"

repeat\_region

repeat\_region /note="14 copies of 2 mer 93 % conserved"

repeat\_region

repeat\_region /note="13 copies of 2 mer 92 % conserved"

repeat\_region

repeat\_region /note="Alu repeat: matches 300. .1 of consensus"

repeat\_region

repeat\_region /note="13 copies of 2 mer 92 % conserved"

repeat\_region

repeat\_region /note="Alu repeat: matches 268. .65 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 1. .306 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 34. .306 of consensus"

repeat\_region

repeat\_region /note="L1 element fragment"

repeat\_region

repeat\_region /note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

repeat\_region /note="L1 element fragment"

repeat\_region

repeat\_region /note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

repeat\_region /note="L1 element fragment"

repeat\_region

repeat\_region /note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 302. .66 of consensus"

repeat\_region

repeat\_region /note="Alu repeat: matches 306. .34 of consensus"

repeat\_region

/note="2 copies of 133 mer 93 % conserved"

repeat\_region

27850. .27930

repeat\_region

/note="L1 element fragment"

repeat\_region

28002. .28043

repeat\_region

/note="L1 element fragment"

repeat\_region

28323. .28780

repeat\_region

/note="L1 element fragment"

repeat\_region

29176. .29449

repeat\_region

/note="Alu repeat: matches 1. .306 of consensus"

repeat\_region

29904. .30192

repeat\_region

/note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

30354. .30663

repeat\_region

/note="L1 element fragment"

repeat\_region

30683. .30972

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

30977. .31204

repeat\_region

/note="L1 element fragment"

repeat\_region

34819. .35121

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

37059. .37282

repeat\_region

/note="MLT2B2 element fragment"

repeat\_region

37453. .37506

repeat\_region

/note="MLT2B2 element fragment"

repeat\_region

38414. .38706

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

39516. .39563

repeat\_region

/note="24 copies of 2 mer 81 % conserved"

repeat\_region

41244. .41535

repeat\_region

/note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

43994. .44242

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 44. .308 of consensus"

repeat\_region

complement(44369. .44677)

misc\_feature

/note="match: 217151 DNA segment containing (CA) repeat"

repeat\_region

complement(44479. .44510)

repeat\_region

/note="16 copies of CA 100% conserved; Polymorphic by comparison with 217151"

repeat\_region

44702. .44889

repeat\_region

/note="2 copies of 94 mer 85 % conserved"

repeat\_region

45935. .46229

repeat\_region

/note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

46910. .46988

repeat\_region

/note="L1 element fragment"

repeat\_region

47081. .47119

repeat\_region

/note="L1 element fragment"

repeat\_region

47168. .47431

repeat\_region

/note="L1 element fragment"

repeat\_region

47615. .47902

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

47922. .48213

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

49426. .49717

repeat\_region

/note="Alu repeat: matches 1. .308 of consensus"

repeat\_region

50248. .50533

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 308. .1 of consensus"

repeat\_region

51097. .51380

repeat\_region

/note="Alu repeat: matches 1. .304 of consensus"

repeat\_region

62318. .62543

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 302. .66 of consensus"

repeat\_region

63358. .63613

repeat\_region

/partial

repeat\_region

/note="Alu repeat: matches 306. .34 of consensus"

repeat\_region

66191. .66236

repeat\_region

/note="23 copies of 2 mer 80 % conserved"

repeat\_region

68771. .69062

repeat\_region

/partial







## ORIGIN

	Query Match	4.48;	Score 232.2;	DB 89;	Length 2212;
	Best Local Similarity	62.1%;	Pred. No. 1.6e-33;		
	Matches 366;	Conservative 0;	Mismatches 223;	Indels 0;	Gaps 0;
3010	aattatgtctctaccagagtatggttccataatacaaticacaccttttgccacatg	3069			
1	AATTATATCACTGTGGCTGAATGATGTAGTTCATACATGATTCACCTGCTAGCCCATG	60			
3070	accagagatttgtcaagtcagggatattgaacaacttaagaagttaaagaatgtcttt	3129			
61	ATCCAGATTTTACAAGATCACAAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCCTAT	120			
3130	ggttgctctggaaataattaatggcctaataaatgaaaaataaacagtcacgccttttatcagaa	3189			
121	GGTTCATGCTTCGAAGTTTTAATGACAAAGAATGAANAACAAATGAGCCATGCCCTTTATGAAGA	180			
3190	agatggttagaaaatataaacaaacaaagaatgcccaaggaccagatgatgcataaaatga	3249			
181	AGATGCGAGAGAACATCAAGTTAACACAGAGATGCCCAGCTCCAGATGAATCCAAGACAA	240			
3250	atgaaaaactgtcacactgtgtgtgtagtggcatgaaatcatcatcatgtccaagagtgacta	3309			
241	ATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAATAAGTGCCTT	300			
3310	catacagtttggaaatctcctaagacccggtactaccagctcggtttcttactcaaacctg	3369			
301	TGTGCATGCAGATTACC AAAGHACCCAGCTCTCCCAATGNAATTTTTTACACAACTG	360			
3370	acaagaatttcagtaacaccaaaaattatctgcctcctgaaatgaatacatttttcactc	3429			
361	AAAAGGACTTCTGTAAACGATAGAGCTTATATTTCAGAAGACAAAGAGTACTTCTGTGTTAA	420			
3430	ctggaaaaccttaaaacaccaaagtcttagagagctgtttaacaagccactttcatacgag	3489			
421	CAGGAAAGCCAAAGCGCTGCTGGAGTACTAGGTGCGAGTAAATAAGCCGCTTATCAGCAACGG	480			
3490	gcaagcaatctcagaccacaaatcatcacgaaatggaaactgtaagcaatgcaagcagcagct	3549			
481	GAAGGNAACCCCTATGTTAGAACCATGGCAGCTGGGAAGCAANTATTATGTAAATTT	540			
3550	caaatcccaagctctcctggaaagaaataaaggggagggttgatgtcttga	3598			
541	CAGAGCTGAACCTCTTCAACCGGAAATCCGATCAAGGGAACAGAGCTTCAAG	589			

**RESULT** 14

AC016449	AC016449	72157 bp	DNA	HTG
LOCUS				
DEFINITION	Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.			
ACCESSION	AC016449			
VERSION	AC016449.2	GI:9104517		
KEYWORDS	HTG; HTGS_PHASE0.			
SOURCE	human.			
ORGANISM	Homo sapiens			

REFERENCE	AUTHORS	TITLE	JOURNAL
1	(bases 1 to 72157)	Blirren, B., Linton, L., Nusbaum, C. and Lander, E.	Primates: Catarrhini, Hominoidea; Homo.
2	(bases 1 to 72157)	Homo sapiens, clone RP11-45L14	Unpublished
3	(bases 1 to 72157)	Blirren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,	

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollard, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Romano, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye-W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6479175.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----

Center project name: L1189

Center clone name: 45 L 14

\* NOTE: This record contains 83 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

757:	contig of 757 bp in length
857:	gap of 100 bp
858:	1635: contig of 778 bp in length
1636:	1735: gap of 100 bp
1736:	2490: contig of 755 bp in length
2491:	2590: gap of 100 bp
2591:	3343: contig of 753 bp in length
3344:	3443: gap of 100 bp
3444:	4224: contig of 781 bp in length
4225:	4324: gap of 100 bp
4325:	5088: contig of 764 bp in length
5089:	5188: gap of 100 bp
5189:	5968: contig of 780 bp in length
5969:	6068: gap of 100 bp
6069:	6844: contig of 776 bp in length
6845:	6944: gap of 100 bp
6945:	7701: contig of 757 bp in length
7702:	7801: gap of 100 bp
7802:	8573: contig of 772 bp in length
8574:	8673: gap of 100 bp
8674:	9430: contig of 757 bp in length
9431:	9530: gap of 100 bp
9531:	10295: contig of 765 bp in length
10296:	10395: gap of 100 bp
10396:	11139: contig of 744 bp in length
11140:	11239: gap of 100 bp
11240:	11988: contig of 749 bp in length
11989:	12088: gap of 100 bp
12089:	12874: contig of 786 bp in length
12875:	12974: gap of 100 bp
12975:	13750: contig of 776 bp in length
13751:	13850: gap of 100 bp
13851:	14638: contig of 788 bp in length
14639:	14738: gap of 100 bp
14739:	15520: contig of 782 bp in length
15521:	15620: gap of 100 bp
15621:	16381: contig of 761 bp in length
16382:	16481: gap of 100 bp
16482:	17248: contig of 767 bp in length
17249:	17348: gap of 100 bp
17349:	18123: contig of 775 bp in length



```
misc_feature 301501..303502
/note="assembly_fragment:01554"
misc_feature 303603..306492
/note="assembly_fragment:02456"
misc_feature 306593..309852
/note="assembly_fragment:03413"
misc_feature 309953..312398
/note="assembly_fragment:03967"
misc_feature 312499..315866
/note="assembly_fragment:04102"
misc_feature 315967..318753
/note="assembly_fragment:04448"
misc_feature 318854..323190
/note="assembly_fragment:04616"
misc_feature 323291..329923
/note="assembly_fragment:04963"
misc_feature 330024..332405
/note="assembly_fragment:06188"
misc_feature 332506..335631
/note="assembly_fragment:06207"
misc_feature 335732..338304
/note="assembly_fragment:06594"
misc_feature 338405..340969
/note="assembly_fragment:06804"
BASE COUNT 100857 a 71110 c 68176 g 96314 t 4512 others
ORIGIN

Query Match 7.6%; Score 399; DB 81; Length 340969;
Best Local Similarity 73.3%; Pred. No. 1.2e-64;
Matches 729; Conservative 0; Mismatches 215; Indels 51; Gaps 15;

QY 4270 tcaaatattttgcacaaagttcctaattgttaaacatacatattgtgtatttaattc 4329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43949 TAAAGATGTTTAAATAATGATGGCTGGTACTAGGCTAAATTCGGATGTTTCACATTG 43890

QY 4330 catattatggcccccattacactaggtacgcgcgcgaagtgcctaaaggaggaacgcgatg 4389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43889 GTCATCTGTTGACTTTTCTCTTCAGGTCACGACGCGGAAGCTCCAAACGAGAGACGATG 43830

QY 4390 acaaatgttaataaactctctctgtgaagctttgtggaaaaatctttttttttttt 4449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43829 AACACGTC-----ATTAACTCTATTGTGAAGCTTTGGGGGAAAAACATTTTGTGTC 43775

QY 4450 tttttttgtcgaagctgaggtgaataaagccttgatgcacaaatggagctgta 4509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43774 T-----AAATATGAGGCTAAATAAGCCTTGATGTCACAAACCAAGCACTGCCAA 43726

QY 4510 agagtgacagtgtagacttactttgtgaccccatcacatttgggtgcacatgctttagc 4569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43725 AGCGTGACAGATTGGAC-----TTCCGATGACCATGGGTTTGGTGCACAGGTATCAGC 43672

QY 4570 catacaca-----tggtaacattgactatggagcttcttgaaagtgttaatgtgcgatgcta 4626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43671 CATCCACACAGTGATAACACTGACTCCAGAGTCTTGTGAAAGTGCTCTGTGTGATGCTG 43612

QY 4627 tgtagacataaagaagaacttgaataatctttttcttttttttttttttttttttttt 4686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43611 TGTAGACACAACTAAAGAGACACATGTAATAATCCCACTCCCCCA--TTTCGTGACTT 43554

QY 4687 ctgaagtgctgtatagcttttttattctgcgcgttttaaaactgacagtaacccagctgttatt 4746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43553 CTAAAGTGCTGTATAGCTTTTATCTCGGGCTTTAAACTGCACAGGACCCAGCTGCTACC 43494

QY 4747 ggaatctattgattgaaagaattgttaggatagatctttaagca----gtaatcttgc 4802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43493 GGATCTCTTATTGAAAGCAATTGTAGGGTGGATCTTTAAGCAGTACGTAGTCTGTCA 43434

QY 4803 gtgtttgattgtatttcttgaatttactgtgaaaaaaatttttttcaacaattg 4862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43433 GTGTTGTATTGTAATTCCTGCAATTTTACTGTG----AAATTTGTTTTCACAAATTG 43378

QY 4863 gtgtcattttctgtatgcactattttgttggag---agttaaatggctctcttccctttgt 4919
```

```
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43377 GTGTCAATTTTCTCAATGTCACACTGTTTCTTGAGAGTTAAATGGTTTCTTCTGCCCTCTTT 43318

QY 4920 gtattctaacctagtgatttaactctctggcacccttaactctcaagagtgctaaatgtctg 4979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43317 GTAGCTCATCTA-TGTTTACTCTCAGGGCACTGTT-GTCTTCTGAGGTGCTAAAGCATCTG 43260

QY 4980 ccattacaccagaagtgctctgataggagacacacatgcacaaattgtgaaatagtc 5039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43259 CCAGCACACTAGAA-CATGGCTCTGTCTGGGCGACACATGG---ATTGTGAGATAGTCC 43204

QY 5040 tgaagttcttgattacttaccactcagtcagtattgttccacagaattttctggccttt 5099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43203 TGAGCTTATCTGATGACTTCACACCTCAGTATTG---GTCCACAGAAATCTCTGCCCTTT 43148

QY 5100 catggcaatgaaatttttaagaagaagatttaagatttttaatttaagaagatgctt 5159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43147 CATGGCAACGAAATTTTAAAGAA-AGAGATTTTAAATTTTAAAGAGTGTGTT 43089

QY 5160 ataaataatgtactgaattcttccatcttccatcttccatctc--cttcagtttttattaa 5217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43088 ATAAATAATGTACTAAATTTCTTTTATTCATTCATTCATCTTTTTCAGTTTATATTAAT 43029

QY 5218 ctactgtatcaataaaaaattctgtaatttgaatgag 5252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43028 CTACTGTATCAATAAAATTTCTGTAATTTGAATGG 42994

RESULT 13
AK021757 2212 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA FLJ11695 fis, clone HEMBA1005019, highly similar
DEFINITION to Homo sapiens mRNA for KIAA0648 protein.
ACCESSION AK021757
VERSION AK021757.1 GI:10433003
KEYWORDS oligo capping; fis (full insert sequence);
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
ORGANISM mRNA, clone_lib:HEMBA1 clone:HEMBA1005019.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nishikawa,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished (2000)
REFERENCE 2 (bases 1 to 2212)
AUTHORS Isogai,T. and Otsuki,T.
DIRECT SUBMISSION
Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
source 1. .2212
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1005019"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/tissue_type="whole embryo, mainly head"
/notes="cloning vector: pME18SFL3"
BASE COUNT 751 a 385 c 430 g 646 t
```



```
DEFINITION Mus musculus chromosome 5 clone RP21-583E8, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL512630
VERSION AL512630.2 GI:13443466
KEYWORDS HTG: HTGS-PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 340969)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
COMMENT On Mar 24, 2001 this sequence version replaced gi:12193249.
----- GenomE Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: dm583E8
----- Summary Statistics
Assembly program: XMAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 323281 bases at least Q40
Consensus quality: 329435 bases at least Q30
Consensus quality: 332334 bases at least Q20
Insert size: 336469; sum-of-contigs
Insert size: 189900; 1.3% error; agarose-fp
Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
coverage: 11.68x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1..340969
Location/Qualifiers
db_xref="taxon:10090"
/chromosome="5"
/clone="RP21-583E8"
/clone_lib="RPCI-21"
1..10854
/note="assembly_fragment:06116
fragment_chain:1"
10955..16884
/note="assembly_fragment:05022
fragment_chain:1"
16985..22344
/note="assembly_fragment:04089
fragment_chain:1"
22445..36631
/note="assembly_fragment:04629
fragment_chain:1"
36732..63941
/note="assembly_fragment:02431
fragment_chain:1"
64042..76959
/note="assembly_fragment:06652
fragment_chain:2"
77060..80361
/note="assembly_fragment:02856
fragment_chain:2"
80462..84581
/note="assembly_fragment:05282
fragment_chain:2"
84682..89169
/note="assembly_fragment:01199
fragment_chain:2"
misc_feature
89270..94454
/note="assembly_fragment:00854
fragment_chain:3"
94555..99344
/note="assembly_fragment:04769
fragment_chain:3"
99445..105364
/note="assembly_fragment:03592
fragment_chain:3"
105465..112857
/note="assembly_fragment:02797
fragment_chain:4"
112958..116353
/note="assembly_fragment:05253
fragment_chain:4"
116454..119029
/note="assembly_fragment:01663
fragment_chain:4"
119130..157992
/note="assembly_fragment:03653
fragment_chain:5"
158093..164601
/note="assembly_fragment:05568
fragment_chain:5"
164702..195901
/note="assembly_fragment:01977
fragment_chain:5"
196002..198807
/note="assembly_fragment:00023
fragment_chain:6"
198908..204604
/note="assembly_fragment:05254
fragment_chain:6"
204705..207838
/note="assembly_fragment:00245
fragment_chain:7"
207939..221335
/note="assembly_fragment:03587
fragment_chain:7"
221436..224577
/note="assembly_fragment:00642
fragment_chain:8"
224678..228248
/note="assembly_fragment:04656
fragment_chain:8"
228349..237923
/note="assembly_fragment:01666
fragment_chain:9"
238024..242134
/note="assembly_fragment:05529
fragment_chain:9"
242235..250341
/note="assembly_fragment:02430
fragment_chain:10"
250442..273524
/note="assembly_fragment:00629
fragment_chain:10"
273625..277125
/note="assembly_fragment:04232
fragment_chain:11"
277226..282694
/note="assembly_fragment:02725
fragment_chain:11"
282795..284888
/note="assembly_fragment:04893
fragment_chain:12"
284989..290372
/note="assembly_fragment:01882
fragment_chain:12"
290473..298809
/note="assembly_fragment:00742"
298910..301400
/note="assembly_fragment:00762"
```







Query Match 8.4%; Score 443; DB 64; Length 72157;  
Best Local Similarity 98.9%; Pident. No. 7.3e-73;  
Matches 446; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

3689 gtctgaattgagaaaccttagaggcagagaataaaccccgctcacagaacagagagaa 3748  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3686 GTCTGAATTGGAGAACCTTAGAGGACAGGAAAAAACCCTCCATCACAAGACAGAGAA 39627  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3749 attagtgtgatgactgacttaagtgtgtatgagaaacccaacctaagggcagtcgcg 3808  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39626 ATTAGGTATGGTACTTGACTTAAGTTGGTACAGGAACAGAACCCTTAAGGCAGTCAGCG 39567  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3809 agtgcgaaaagagccatcacgcttcagaaatctgatgaacagcagtgccctgaggaata 3868  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39566 AAGTCGAAAAGAGGCCATACCGCTTCAGAACTCTGATGAACAGCAGTCGCTCAGGAAAA 39507  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3869 gaggctcaaaagaatatattagaataatgaagataacagaataatgccgcaaaaaagg 3928  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39506 GAGGCTCAAGAAGATATATTAGAAATYGAAGATGAACAGATAGTCCGCCAAAAAAGGG 39447  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3929 taaaagagggcgcaccaccaaacctcttgtgtgaggtacaccaagaagagccacaat 3988  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39446 TAAAGAGGCCGACCAACCAACCTCTTGTGTGAGGTACACCAAGAAGAGCCACAT 39387  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3989 gaaactcttaaaaagaagcaaaaaaatctggactccagcaccagagagagga 4048  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39386 GAAAACTTCTTAAAAAAGGAAGCAAAAAAATCTGACCTCCAGCACAGAGGAGGAGGA 39327  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
4049 agaagaagaagacaagtggaaatacgaacagaagtcgaagtcgaagcaaacagcaccgagt 4108  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39326 AGAAGAAGAAGACAAAGTGGAAATACGGAACAGAAAGTCCAAGAACCAACAGACCCGAGT 39267  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
4109 gtcaaaggagacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4139  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
39266 GTCAAGGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 39236  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11  
XCUS 068224/c

DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS SEQUENCE SAMPLING.

AC068224 AC068224.1 GI:7671284 HTG 30-APR-2000

WORDS HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 54398)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 3, clone RP11-660H19

Unpublished

2 (bases 1 to 54398)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goette,M., Graham,L., Grand-Pierre,N., Grant,G., Haqos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Menues,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,



\* 1 757: contig of 757 bp in length  
\* 758 857: gap of 100 bp  
\* 858 1635: contig of 778 bp in length  
\* 1636 1735: gap of 100 bp  
\* 1736 2490: contig of 755 bp in length  
\* 2491 2590: gap of 100 bp  
\* 2591 3343: contig of 753 bp in length  
\* 3344 3443: gap of 100 bp  
\* 3444 4224: contig of 781 bp in length  
\* 4225 4324: gap of 100 bp  
\* 4325 5088: contig of 764 bp in length  
\* 5089 5188: gap of 100 bp  
\* 5189 5968: contig of 780 bp in length  
\* 5969 6068: gap of 100 bp  
\* 6069 6844: contig of 776 bp in length  
\* 6845 6944: gap of 100 bp  
\* 6945 7701: contig of 757 bp in length  
\* 7702 7801: gap of 100 bp  
\* 7802 8573: contig of 772 bp in length  
\* 8574 8673: gap of 100 bp  
\* 8674 9430: contig of 757 bp in length  
\* 9431 9530: gap of 100 bp  
\* 9531 10295: contig of 765 bp in length  
\* 10296 10395: gap of 100 bp  
\* 10396 11139: contig of 744 bp in length  
\* 11140 11239: gap of 100 bp  
\* 11240 11988: contig of 749 bp in length  
\* 11989 12088: gap of 100 bp  
\* 12089 12874: contig of 786 bp in length  
\* 12875 12974: gap of 100 bp  
\* 12975 13750: contig of 776 bp in length  
\* 13751 13850: gap of 100 bp  
\* 13851 14638: contig of 788 bp in length  
\* 14639 14738: gap of 100 bp  
\* 14739 15520: contig of 782 bp in length  
\* 15521 15620: gap of 100 bp  
\* 15621 16381: contig of 761 bp in length  
\* 16382 16481: gap of 100 bp  
\* 16482 17248: contig of 767 bp in length  
\* 17249 17348: gap of 100 bp  
\* 17349 18123: contig of 775 bp in length  
\* 18124 18223: gap of 100 bp  
\* 18224 18977: contig of 754 bp in length  
\* 18978 19077: gap of 100 bp  
\* 19078 19860: contig of 783 bp in length  
\* 19861 19960: gap of 100 bp  
\* 19961 20732: contig of 772 bp in length  
\* 20733 20832: gap of 100 bp  
\* 20833 21590: contig of 758 bp in length  
\* 21591 21690: gap of 100 bp  
\* 21691 22446: contig of 756 bp in length  
\* 22447 22546: gap of 100 bp  
\* 22547 23333: contig of 787 bp in length  
\* 23334 23433: gap of 100 bp  
\* 23434 24212: contig of 779 bp in length  
\* 24213 24312: gap of 100 bp  
\* 24313 25093: contig of 781 bp in length  
\* 25094 25193: gap of 100 bp  
\* 25194 25958: contig of 765 bp in length  
\* 25959 26058: gap of 100 bp  
\* 26059 26839: contig of 781 bp in length  
\* 26840 26939: gap of 100 bp  
\* 26940 27705: contig of 766 bp in length  
\* 27706 27805: gap of 100 bp  
\* 27806 28568: contig of 763 bp in length  
\* 28569 28668: gap of 100 bp  
\* 28669 29419: contig of 751 bp in length  
\* 29420 29519: gap of 100 bp  
\* 29520 30311: contig of 792 bp in length  
\* 30312 30411: gap of 100 bp  
\* 30412 31202: contig of 791 bp in length  
\* 31203 31302: gap of 100 bp  
\* 31303 32082: contig of 780 bp in length

\* 32083 32182: gap of 100 bp  
\* 32183 32934: contig of 752 bp in length  
\* 32935 33034: gap of 100 bp  
\* 33035 33796: contig of 762 bp in length  
\* 33797 33896: gap of 100 bp  
\* 33897 34683: contig of 787 bp in length  
\* 34684 34783: gap of 100 bp  
\* 34784 35552: contig of 769 bp in length  
\* 35553 35652: gap of 100 bp  
\* 35653 36406: contig of 754 bp in length  
\* 36407 36506: gap of 100 bp  
\* 36507 37294: contig of 788 bp in length  
\* 37295 37394: gap of 100 bp  
\* 37395 38159: contig of 765 bp in length  
\* 38160 38259: gap of 100 bp  
\* 38260 39017: contig of 758 bp in length  
\* 39018 39117: gap of 100 bp  
\* 39118 39890: contig of 773 bp in length  
\* 39891 39990: gap of 100 bp  
\* 39991 40774: contig of 784 bp in length  
\* 40775 40874: gap of 100 bp  
\* 40875 41643: contig of 769 bp in length  
\* 41644 41743: gap of 100 bp  
\* 41744 42521: contig of 778 bp in length  
\* 42522 42621: gap of 100 bp  
\* 42622 43393: contig of 772 bp in length  
\* 43394 43493: gap of 100 bp  
\* 43494 44288: contig of 775 bp in length  
\* 44289 44388: gap of 100 bp  
\* 44389 45149: contig of 781 bp in length  
\* 45150 45249: gap of 100 bp  
\* 45250 46024: contig of 775 bp in length  
\* 46025 46124: gap of 100 bp  
\* 46125 46868: contig of 744 bp in length  
\* 46869 46968: gap of 100 bp  
\* 46969 47720: contig of 752 bp in length  
\* 47721 47820: gap of 100 bp  
\* 47821 48596: contig of 776 bp in length  
\* 48597 48696: gap of 100 bp  
\* 48697 49462: contig of 766 bp in length  
\* 49463 49562: gap of 100 bp  
\* 49563 50335: contig of 773 bp in length  
\* 50336 50435: gap of 100 bp  
\* 50436 51207: contig of 772 bp in length  
\* 51208 51307: gap of 100 bp  
\* 51308 52078: contig of 771 bp in length  
\* 52079 52178: gap of 100 bp  
\* 52179 52954: contig of 776 bp in length  
\* 52955 53054: gap of 100 bp  
\* 53055 53813: contig of 759 bp in length  
\* 53814 53913: gap of 100 bp  
\* 53914 54675: contig of 762 bp in length  
\* 54676 54775: gap of 100 bp  
\* 54776 55537: contig of 762 bp in length  
\* 55538 55637: gap of 100 bp  
\* 55638 56413: contig of 776 bp in length  
\* 56414 56513: gap of 100 bp  
\* 56514 57278: contig of 765 bp in length  
\* 57279 57378: gap of 100 bp  
\* 57379 58149: contig of 771 bp in length  
\* 58150 58249: gap of 100 bp  
\* 58250 59023: contig of 774 bp in length  
\* 59024 59123: gap of 100 bp  
\* 59124 59906: contig of 783 bp in length  
\* 59907 60006: gap of 100 bp  
\* 60007 60792: contig of 786 bp in length  
\* 60793 60892: gap of 100 bp  
\* 60893 61663: contig of 771 bp in length  
\* 61664 61763: gap of 100 bp  
\* 61764 62539: contig of 776 bp in length  
\* 62540 62639: gap of 100 bp  
\* 62640 63422: contig of 783 bp in length



* 38331	38430: gap of	100 bp
* 38431	39109: contig of 679 bp	in length
* 39110	39209: gap of	100 bp
* 39210	39912: contig of 703 bp	in length
* 39913	40012: gap of	100 bp
* 40013	40712: contig of 700 bp	in length
* 40713	40812: gap of	100 bp
* 40813	41514: contig of 702 bp	in length
* 41515	41614: gap of	100 bp
* 41615	42294: contig of 680 bp	in length
* 42295	42394: gap of	100 bp
* 42395	43094: contig of 700 bp	in length
* 43095	43194: gap of	100 bp
* 43195	43915: contig of 721 bp	in length
* 43916	44015: gap of	100 bp
* 44016	44740: contig of 725 bp	in length
* 44741	44840: gap of	100 bp
* 44841	45533: contig of 693 bp	in length
* 45534	45633: gap of	100 bp
* 45634	46330: contig of 697 bp	in length
* 46331	46430: gap of	100 bp
* 46431	47137: contig of 707 bp	in length
* 47138	47237: gap of	100 bp
* 47238	47927: contig of 690 bp	in length
* 47928	48027: gap of	100 bp
* 48028	48709: contig of 682 bp	in length
* 48710	48809: gap of	100 bp
* 48810	49509: contig of 700 bp	in length
* 49510	49609: gap of	100 bp
* 49610	50331: contig of 722 bp	in length
* 50332	50431: gap of	100 bp
* 50432	51135: contig of 704 bp	in length
* 51136	51235: gap of	100 bp
* 51236	51981: contig of 746 bp	in length
* 51982	52081: gap of	100 bp
* 52082	52783: contig of 702 bp	in length
* 52784	52883: gap of	100 bp
* 52884	53580: contig of 697 bp	in length
* 53581	53680: gap of	100 bp
* 53681	54398: contig of 718 bp	in length.
	Location/Qualifiers	
	1..54398	

Query Match	9.4%	Score 497.4	DB 73	Length 54398
Best Local Similarity	92.7%	Pred. No. 5.2e-83		
Matches 522	Conservative 0	Mismatches 41	Indels 0	Gaps
QY 4305	catacatattgctgtatttaaatccatatatttagccccattacactaggtacggcgcc	4364		
Db 32078	CCTGCACATATACCATTTTTAATTAAGCATCTGGTGACTTTCCTTTTAAAGGTACGGCGC	32137		
QY 4365	gaagtgtctaaagggagcgcatgaacaaatgtaattaaataactttctctgtgaaagct	4424		
Db 32138	GAAGTGTCTAAAGGGAGCGGCATGAACAAATGTAATTAATTAACCTTCTCTGTGAAGCT	32197		
QY 4425	ttggaataactcttt	4484		
Db 32198	TTGAAATAACTTTTTTTTTTTTTTTTTTTTTTTTGGGCAAGCTTGAGGCTGAATAAAGCT	32257		
QY 4485	ttgatgcacaaaaatgggaactctgaagagtgagacagtttgagaccttacttttgtagaccca	4544		
Db 32258	TTGATGCACAAAATGGGAGCTGCTGAAGAGTGGACAGTTGGACCTTACTTTGGTGACCCCA	32317		
QY 4545	tacatttggtgcacatgcttttagcacaacacatggtaacattgactatgagctctgtg	4604		
Db 32318	TACATTTGTGTGCATGCTTTTAGCCATACACATGGGTAAACATTGACTATGGAGTCTTG	32377		
QY 4605	aaagtgtaatgtcgatggctatgtagacataaagaagaacttgtaaatatctctttttc	4664		
Db 32378	AAAGTGTAAATGTGGATGGCTATGTAGACATTAAGAAGAAACTTGGAAATATCTTTTTTC	32437		
QY 4665	tttttttaattgtttctgattctgaagtgtctgtatagcttttatctgcggcctttaaac	4724		

Db	32438	TTTTTTTTTAATGGTCTCGATTTTCGGAAGTGGTTGTATAGCTTTTATCTCGCGCTTTAAAC	32497
Qy	4725	tgacgtagccgcgaactgtttattggatcattgattgaaagaattttgttaggatagatc	4784
Db	32498	TGACAGTAGCCGACTGTTTATTATGGATCTATTGATTTGAAAAGAAATTTCTTAGGATAGATC	32557
Qy	4785	ttaagcagtaactctgcagttgttattgtattttctcaattttactctgaaaaaaa	4844
Db	32558	TTAAGCAGTAATCTGTCACTGTTGTATTGTCTTGTGTGGAAATTTTCTCTGGAATAAA	32617
Qy	4845	attgtttttcaacaattggctc	4867
Db	32618	ATTTGTTCTCAACAATCGGTGCC	32640

RESULT	10	AC016449	72157 bp	DNA	HTG	13-JUL-2000
LOCUS	AC016449/c					
DEFINITION		Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC016449					
VERSION	AC016449.2	GI:9104517				
KEYWORDS	HTG; HTGS_PHASE0.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 72157)				
TITLE		Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 72157)				
AUTHORS		Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,R., Donelan,L., Doyle,M., Ferrelar,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
TITLE		Direct Submission				
JOURNAL		Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT		On Jul 13, 2000 this sequence version replaced gi:6479175. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>				
		----- Genome Center				
		Center: Whitehead Institute/ MIT Center for Genome Research				
		Center code: WtBR				
		Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a>				
		Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a>				
		----- Project Information				
		Center project name: L1189				
		Center clone name: 45_L14				
		-----				
		* NOTE: This record contains 83 individual				
		* sequencing reads that have not been assembled into				
		* contigs. Runs of N are used to separate the reads				
		* and the order in which they appear is completely				
		* arbitrary. Low-pass sequence sampling is useful for				
		* identifying clones that may be gene-rich and allows				
		* overlap relationships among clones to be deduced.				
		* However, it should not be assumed that this clone				
		* will be sequenced to completion. In the event that				
		* the record is updated, the accession number will				
		* be preserved.				



AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Homo sapiens chromosome 3, clone RP11-660H19  
Unpublished  
2 (bases 1 to 54398)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,  
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
Galaqan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meldrum, J., Menues, L., Mhova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WBIR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10161  
Center clone name: 660\_H\_19  
-----

\* NOTE: This record contains 68 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be generic and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 685: contig of 685 bp in length  
\* 686 785: gap of 100 bp  
\* 786 1513: contig of 728 bp in length  
\* 1514 1613: gap of 100 bp  
\* 1614 2354: contig of 741 bp in length  
\* 2355 2454: gap of 100 bp  
\* 2455 3184: contig of 730 bp in length  
\* 3185 3284: gap of 100 bp  
\* 3285 3959: contig of 675 bp in length  
\* 3960 4059: gap of 100 bp  
\* 4060 4746: contig of 687 bp in length  
\* 4747 4846: gap of 100 bp  
\* 4847 5564: contig of 718 bp in length  
\* 5565 5664: gap of 100 bp  
\* 5665 6360: contig of 696 bp in length  
\* 6361 6460: gap of 100 bp  
\* 6461 7132: contig of 672 bp in length  
\* 7133 7232: gap of 100 bp  
\* 7233 7937: contig of 705 bp in length  
\* 7938 8037: gap of 100 bp  
\* 8038 8749: contig of 712 bp in length  
\* 8750 8849: gap of 100 bp  
\* 8850 9522: contig of 673 bp in length  
\* 9523 9622: gap of 100 bp  
\* 9623 10356: contig of 734 bp in length  
\* 10357 10456: gap of 100 bp  
\* 10457 11181: contig of 725 bp in length  
\* 11182 11281: gap of 100 bp  
\* 11282 11975: contig of 694 bp in length  
\* 11976 12075: gap of 100 bp  
\* 12076 12785: contig of 710 bp in length  
\* 12786 12885: gap of 100 bp  
\* 12886 13557: contig of 672 bp in length  
\* 13558 13657: gap of 100 bp  
\* 13658 14347: contig of 690 bp in length  
\* 14348 14447: gap of 100 bp  
\* 14448 15132: contig of 685 bp in length  
\* 15133 15232: gap of 100 bp  
\* 15233 15912: contig of 680 bp in length  
\* 15913 16012: gap of 100 bp  
\* 16013 16731: contig of 719 bp in length  
\* 16732 16831: gap of 100 bp  
\* 16832 17506: contig of 675 bp in length  
\* 17507 17606: gap of 100 bp  
\* 17607 18339: contig of 733 bp in length  
\* 18340 18439: gap of 100 bp  
\* 18440 19109: contig of 670 bp in length  
\* 19110 19209: gap of 100 bp  
\* 19210 19900: contig of 691 bp in length  
\* 19901 20000: gap of 100 bp  
\* 20001 20693: contig of 693 bp in length  
\* 20694 20793: gap of 100 bp  
\* 20794 21464: contig of 671 bp in length  
\* 21465 21564: gap of 100 bp  
\* 21565 22245: contig of 681 bp in length  
\* 22246 22345: gap of 100 bp  
\* 22346 23049: contig of 704 bp in length  
\* 23050 23149: gap of 100 bp  
\* 23150 23857: contig of 708 bp in length  
\* 23858 23957: gap of 100 bp  
\* 23958 24685: contig of 728 bp in length  
\* 24686 24785: gap of 100 bp  
\* 24786 25499: contig of 714 bp in length  
\* 25500 25599: gap of 100 bp  
\* 25600 26302: contig of 703 bp in length  
\* 26303 26402: gap of 100 bp  
\* 26403 27108: contig of 706 bp in length  
\* 27109 27208: gap of 100 bp  
\* 27209 27910: contig of 702 bp in length  
\* 27911 28010: gap of 100 bp  
\* 28011 28720: contig of 710 bp in length  
\* 28721 28820: gap of 100 bp  
\* 28821 29484: contig of 664 bp in length  
\* 29485 29584: gap of 100 bp  
\* 29585 30328: contig of 744 bp in length  
\* 30329 30428: gap of 100 bp  
\* 30429 31111: contig of 683 bp in length  
\* 31112 31211: gap of 100 bp  
\* 31212 31931: contig of 720 bp in length  
\* 31932 32031: gap of 100 bp  
\* 32032 32710: contig of 679 bp in length  
\* 32711 32810: gap of 100 bp  
\* 32811 33502: contig of 692 bp in length  
\* 33503 33602: gap of 100 bp  
\* 33603 34301: contig of 699 bp in length  
\* 34302 34401: gap of 100 bp  
\* 34402 35100: contig of 699 bp in length  
\* 35101 35200: gap of 100 bp  
\* 35201 35914: contig of 714 bp in length  
\* 35915 36014: gap of 100 bp  
\* 36015 36708: contig of 695 bp in length  
\* 36710 36809: gap of 100 bp  
\* 36810 37518: contig of 709 bp in length  
\* 37519 37618: gap of 100 bp  
\* 37619 38330: contig of 712 bp in length



```

QY 5026 ttgtgaataagtcgaagttcttggaattactttacacctcagtagtgattgtccccaga 5085
|||||
Db 26154 TTGTGAATAAGTCCTGAAGTCTCTGGATTACTTTACACCTCAGTATTGATTGTGCCAGA 26213
|||||
QY 5086 attttctggccttcattcagcaatgaaattttaaagaagaagattttaaagtatttaatt 5145
|||||
Db 26214 ATTTCTGGCCTTCTATCGCAATCAAAATTTTAAAGAAAGAAATTTAAAGTATTTTAAAT 26273
|||||
QY 5146 ttaaagtgctgtataaaataagctactgaattcttcttccattttatccatctctca 5205
|||||
Db 26274 TTAAGAGTGTGTATATAATATGACTGCAATCTTTATCCCATTTTATCATCTTTCA 26333
|||||
QY 5206 gtttttattactactgtatcaataaaaaattctgttaatttgaatgagtaaaaaa 5265
|||||
Db 26334 GTTTTATTAACTACTGTATCAATAAAATCTGTAAATTTGAATGAGTTTAAATAGTCT 26393
|||||
QY 5266 aaaa 5269
|||||
Db 26394 AGAA 26397

RESULT 8
AK026889 2079 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ23236 fis, clone COL00725.
ACCESSION AK026889
VERSION AK026889.1 GI:10439854
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL00725.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2079)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Source Location/Qualifiers
1..2079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL00725"
/clone_lib="COL"
/tissue_type="colon"
/notes="cloning vector pME18SFL3"
304..681
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15584.1"
/db_xref="GI:10439855"
/translaton="MDDLTKLVQEQPKGSQSRKRGHTASEDSDEQWPEEKRLKEDI
LNEEDONPPPKKGRKRPKPPLGGTTPKEPTMTSKSKKKSGPPAPEEEEEER
OSGNTQKSKQKQHRVSRRAQR"
BASE COUNT 759 a 354 c 339 g 627 t

```

## ORIGIN

```

Query Match 10.7%; Score 566.4; DB 89; Length 2079;
Best Local Similarity 87.7%; Pred. No. 7.1e-96;
Matches 670; Conservative 0; Mismatches 1; Indels 93; Gaps 1;

QY 3689 gtctgaattggagaagcctagagcaggaagaaaccccgctcacagaacagaggagaa 3748
|||||
Db 237 GTCTGAATTGAGAGCCTAGAGCGCAGGAGAAACCCCGTCACAGAACAGGAGGAGAA 296
|||||
QY 3749 attaggtatgagtgaactgaattggtacaggaacagaaacccctaaagggcagtcagcg 3808
|||||
Db 297 ATTAGGTATGAGTGACTTGACTAAGTTGGTACAGGAACAGAAACCTTAAGGCGATCAGCG 356
|||||
QY 3809 aagtcggaagaaagggccatactcagactctcagaaatctgtgaacacagcagtggtcctgagaaaa 3868
|||||
Db 357 AGTCGGAAGAAAGGCGCATACGGCTTCAGAACTCTGTATGACACAGCATGGCTGGCTGAGGAAAA 416
|||||
QY 3869 gaggtctcaagaagatatattagaataatgaagatgaacagaatagtcgcgcacaaagggg 3928
|||||
Db 417 GAGGCTCAAGAAAGATATATTAGAAAAATGAAGATGAACAGATAGTCCGCCAAAAAGGG 476
|||||
QY 3929 taaaagagggccaccacccaaacccctcttggtggaggtacacccaaagaaagagccacaat 3988
|||||
Db 477 TAAAGAGGCGCGCACCAACCAAAACCTCTTGTGGAGGTACACCAAAAGAAAGAGCCACAAT 536
|||||
QY 3989 gaaacctcttaaaaaaggaagcaaaaaaaatctggacctccagcaccagagaggagga 4048
|||||
Db 537 GAAACTCTTAAAAAGGAAGCAAAAAAAATCTTGGACCTCCAGCACCACAGAGGAGGAGGA 596
|||||
QY 4049 agaagaagaaagcaagtggaataatcggaacagagagtcgcaaaagcaaacagcaccagtg 4108
|||||
Db 597 AGAAGAAGAAAGACAAAGTGGAAATACGGAAACAGAAAGTCCAAAGCAACACGACCCAGT 656
|||||
QY 4109 gtcaagagagacacagcag----- 4127
|||||
Db 657 GTCAAGGAGAGCAGCAGCAGAGGTAAAGCATGTGTAACCTCTAAACTGTCATCTGTTTCGTAC 716
|||||
QY 4128 -----agagcaga 4135
|||||
Db 717 TATATTATAAATCATAATTTGATGCTATCCACATTTGGGTCTTCCCAAGACAGAGCAGA 776
|||||
QY 4136 atctctgaatctagtgcaattgaatccacacagtcacacccacagaaagagaggaag 4195
|||||
Db 777 ATCTCTGAATCTAGTGCATTTGAATCCACACAGTCCACACACAGAGGAGGAGGAG 836
|||||
QY 4196 accatcaaaacgccatccatccatcaacacaaacaaataatgttaagttaataattac 4255
|||||
Db 837 ACCATCAAAACGCCCATCACCATCAACCAACCAAAAAAATGTGTAGTTGTAATATTAC 896
|||||
QY 4256 attcaaaccaatttcaaatattttgcaaaagtctcctaatttgaacatacatattg 4315
|||||
Db 897 ATTCAACCAATTTCAAAATATTATTGCAAAAGTTCTTAATTTGTAACATACATATTG 956
|||||
QY 4316 ctgtattaaattccatatatttagcccatattacactaggtacg 4359
|||||
Db 957 CTGTATTAAATTCATATATTTAGCCCATTTACACTAGGTAAG 1000
|||||

RESULT 9
AC068224 54398 bp DNA HTG 30-APR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC068224
VERSION AC068224.1 GI:7671284
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 54398)

```







```

72405; Paired with EST T51543 matching this clone; match
H91918"
repeat_region      28327..28360
                    /note="17 copies of 2 mer 88 % conserved"
prim_transcript    <28864..>25013
                    /note="match: 3' EST T51543 clone 72405; Paired with EST
                    T51695 matching this clone"
repeat_region      29300..29456
                    /note="AluJb repeat: matches 146..1 of consensus"
repeat_region      30648..30934
                    /note="AluX repeat: matches 7..302 of consensus"
repeat_region      33156..33439
                    /note="AluX repeat: matches 301..2 of consensus"
repeat_region      34900..35082
                    /note="L1 repeat: matches 425..263 of consensus"
repeat_region      34982..35061
                    /note="L1 repeat: matches 343..264 of consensus"
repeat_region      37923..38245
                    /note="AluX repeat: matches 294..1 of consensus"
repeat_region      38457..38536
                    /note="L1 repeat: matches 61..161 of consensus"
repeat_region      39265..39568
                    /note="L1 repeat: matches 302..3 of consensus"
repeat_region      40952..41046
                    /note="L1 repeat: matches 29..125 of consensus"
repeat_region      43042..43331
                    /note="L1 repeat: matches 301..11 of consensus"
repeat_region      43583..43669
                    /note="L1 repeat: matches 10..99 of consensus"
repeat_region      43881..44083
                    /note="L1 repeat: matches 248..28 of consensus"
repeat_region      44772..45170
                    /note="L1 repeat: matches 388..1 of consensus"
repeat_region      45848..45949
                    /note="L1 repeat: matches 144..41 of consensus"
repeat_region      46073..46373
                    /note="L1 repeat: matches 298..1 of consensus"
repeat_region      49357..49436
                    /note="L1 repeat: matches 198..114 of consensus"
repeat_region      50023..50324
                    /note="L1 repeat: matches 1..302 of consensus"
repeat_region      50826..50892
                    /note="L1 repeat: matches 853..921 of consensus"
repeat_region      51407..51534
                    /note="L1 repeat: matches 4869..4999 of consensus"
repeat_region      51563..51863
                    /note="L1 repeat: matches 1..301 of consensus"
repeat_region      51869..52077
                    /note="L1 repeat: matches 86..298 of consensus;
                    incomplete repeat"
repeat_region      52311..52712
                    /note="L1 repeat: matches 44..475 of consensus"
repeat_region      52698..52781
                    /note="L1 repeat: matches 569..653 of consensus"
repeat_region      52719..52833
                    /note="L1 repeat: matches 591..707 of consensus"
repeat_region      53752..54490
                    /note="L1 repeat: matches 4668..5390 of consensus"
repeat_region      54313..54870
                    /note="L1 repeat: matches 1..540 of consensus"
repeat_region      55008..55520
                    /note="L1 repeat: matches 530..1055 of consensus"
repeat_region      57267..57487
                    /note="L1 repeat: matches 79..298 of consensus;
                    incomplete repeat"
repeat_region      57510..57576
                    /note="L1 repeat: matches 76..143 of consensus"
repeat_region      58670..58699
                    /note="L1 repeat: matches 15 copies of 2 mer 87 % conserved"
repeat_region      58676..58699
                    /note="L1 repeat: matches 15 copies of 2 mer 87 % conserved"
repeat_region      59808..60115
                    /note="L1 repeat: matches 895..571 of consensus"

```

```

60135..60526
/note="L1 repeat: matches 505..108 of consensus"
repeat_region      60531..60815
                    /note="L1 repeat: matches 2..290 of consensus"
repeat_region      60816..60915
                    /note="L1 repeat: matches 114..3 of consensus"
repeat_region      60918..61113
                    /note="L1 repeat: matches 529..320 of consensus"
repeat_region      61115..61379
                    /note="L1 repeat: matches 299..31 of consensus;
                    incomplete repeat"
repeat_region      61380..61675
                    /note="L1 repeat: matches 309..1 of consensus"
repeat_region      61540..61805
                    /note="L1 repeat: matches 5389..5107 of consensus"
repeat_region      61916..62212
                    /note="L1 repeat: matches 1..299 of consensus"
repeat_region      64361..64420
                    /note="L1 repeat: matches 63..1 of consensus"
repeat_region      64816..65101
                    /note="L1 repeat: matches 2..302 of consensus"
repeat_region      65202..65856
                    /note="L1 repeat: matches 902..217 of consensus"
repeat_region      66393..66470
                    /note="L1 repeat: matches 2101..2178 of consensus"
repeat_region      66530..66868
                    /note="L1 repeat: matches 2265..2610 of consensus"
repeat_region      66881..67104
                    /note="L1 repeat: matches 2..230 of consensus"
repeat_region      67104..67855
                    /note="L1 repeat: matches 2613..3397 of consensus"
repeat_region      67999..68677
                    /note="L1 repeat: matches 4255..4942 of consensus"
repeat_region      68693..68972
                    /note="L1 repeat: matches 299..1 of consensus"
repeat_region      68983..69436
                    /note="L1 repeat: matches 4954..5390 of consensus"
repeat_region      69283..70242
                    /note="L1 repeat: matches 4954..5390 of consensus"

Query Match      17.1%; Score 901.6; DB 92; Length 137246;
Best Local Similarity 96.0%; Pred. No. 1.9e-158;
Matches 925; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 4306 atacatatgtgtatttaataatccatatttagccccattacactagtagtcgagcggcg 4365
Db 25955 ATGCACATATACCCATTTTAATTAAGCATCTGTGCTTTTCCCTTTTAAGGTACGGCGG 26014
QY 4366 aagtgctaaaagggaacgagatgaacaaatgaataataacttctctgtgaaagcctt 4425
Db 26015 AAGTGTCTAAAGGGAACGGGATGAACAATGAATTAATAACTTTTCTCTGTGAAAGCTT 26074
QY 4426 tggaaaaatcttttttttttttttttttttttttttttttttttttttttttttttt 4485
Db 26075 TGGAAAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 26134
QY 4486 ttagcacaaaaatggagactgctgaaagtgagcagtgagcacttacttttggtagcccat 4545
Db 26135 TGATGCACAAAAATGGGACTGCTGAAGATGGACAGTTGGACCTTACTTTTGGTGACCCCAT 26194
QY 4546 acatttgtgtcacatgcttttagccatcacatgagtggaacattgagctgtgtgaa 4605
Db 26195 ACATTGTGTGTACATGCTTTAGCCATACACATGGTAACATTGACATGATGAGCTTTGTGA 26254
QY 4606 aagtgtaattgtcgatggctgtatgtacataaagaagaacttggtaaatctttttct 4665
Db 26255 AAGTGAATGTGGATGGCTATGTACACATAAAGAAACATTTGTAATATCTTTTCTTCT 26314
QY 4666 ttttttaattgttctgtatttctgaagtgtgtatagcttttttatctgcggctttaaact 4725
Db 26315 TTTTCTTAATGTTCTGATTTCTGAAAGTGTGATAGCTTTTATCTGCGGCTTTAAACT 26374
QY 4726 gacagtaccgactgtttattgatttattgatttgaagaatttggtaggatact 4785

```



Db	1321	ACCAAGTAAGGCAGATATTTGGCTCAGAGCTGCATAAGGCACCTTGTGAAGTTACTGCTCC	1380
Qy	2872	cacttgatgatggaactctgtgcccttcttgcaaaagatcctgtgaagagagaag	2931
Db	1381	CATTGGAGTATATGGGATCTTTGGCTGTGTGCCAAAGATCTCTGTGAAGGAGAGAG	1440
Qy	2932	ctcatgctaggcaatgttttggtgaaaaataataatgttaagcgggagtcctcgaagcgc	2991
Db	1441	CACACGCACGACAATGTTTACTGAAATAATACGATATACGACGGGAATACATTAAGCAGA	1500
Qy	2992	atgcagctgttagtgaataattatgtctctctaccagatgtattgtttccatatcaaa	3051
Db	1501	ATCCTATGCGTACTGAGAAATTTATATCACTTTGCCCTGGAATATAGTTCATACATGA	1560
Qy	3052	ttcaccttttggcacatgaccagattatgtcaaatgacagattatgaacaacttaag	3111
Db	1561	TTCACTTGCTAGCCATGATCCAGATTTTACAAGATCACAGATGTTGATCAGCTTCGTG	1620
Qy	3112	atgtttaaagaatgtcttcttggtttggtaataattaatggctaaataataaataa	3171
Db	1621	ATATCAAGAGAGTGCCTATGCTTGAAGTTTGAAGTTTAAATGACAAAGAAATGAAACAATA	1680
Qy	3172	gtcacctttttatcaagaagatggttagaataatttaacaacaacaaagaatgcccaagagc	3231
Db	1681	GCCATGCTTTTATGAAGAGATGGCAGAGAAATCAAGTTAACCCAGAGATGCCAGTCTC	1740
Qy	3232	cagatgatgcaaaaatgaatgaataactgtacactgtgtgtgattgtccatgaatatca	3291
Db	1741	CAGATGAATCCAGACAAATGAANAACGTGTATACAGTATGTGATGTGGCTCTCTGTGTTA	1800
Qy	3292	tcatttcaagagtgactacatacagtttggatctctctaaagacccggtactaccagctc	3351
Db	1801	TAAATAGTAAAGTGTGTGTGCAATGTCAGATTCACCAAAAGGACCCAGTCTCCCAATGA	1860
Qy	3352	gttttctactcaactgcagaagaatttcagtaacacacaaataattatctgcctcctgaaa	3411
Db	1861	AAATTTTTACAAACCTGAAAAGGACTTCTGTAACGATAAAGAGTTATATTTTCAGAAGAGA	1920
Qy	3412	tgaatacattttcactcctctggaacacccctaaacacaaatgttctagagctgttaaca	3471
Db	1921	CAGAGTACTTCTGTTAAACAGGAACCCAAAGCCCTGCTGGAGTACTAGTGCAGTAATA	1980
Qy	3472	agcaatttcatcagcaggcaagcaatctcagacaaatcatcacgaatggaatactgaa	3531
Db	1981	AGCCTTTATCAGCAACGGAAGAAACCCCTATGTTAGAAGCAGCTGGCAGTGCAGACTGAA	2040
Qy	3532	gcaatcagcagcagcagctcaaatccaagctctcctctggaagaataaaagggggcttgata	3591
Db	2041	GCAATATTATGTAAATTCAGAGCTGAAACCCCTCAACCGGAAATCGATCAAGGGAACAGA	2100
Qy	3592	gtctgaaatgatcagtcagtgaaaatgaaga	3622
Db	2101	GTTACAGGCAGCAGAACTGGAGTTAGTGA	2131
RESULT	6		
LOCUS	HS49J10		
DEFINITION	Human DNA sequence from PAC 49J10, BRCA2 gene region chromosome 13q12-13 contains ESTs.		
ACCESSION	284572		
VERSION	13q12.1		
KEYWORDS	GI:1813974		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Whiteley, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire, CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		



```

REFERENCE
AUTHORS      Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
TITLE        Direct Submission
JOURNAL      Submitted (26-MAY-1998) to the DDBJ/EMBL/Genbank databases. Osamu
              Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
              Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
              (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
              Fax:+81-438-52-3914)
REFERENCE
AUTHORS      Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
              Kotani,H., Nomura,N. and Ohara,O.
TITLE        Prediction of the coding sequences of unidentified human genes. X.
              The complete sequences of 100 new cDNA clones from brain which can
              code for large proteins in vitro
JOURNAL      DNA Res. 5 (3), 169-176 (1998)
MEDLINE      98403880
FEATURES
source       Location/Qualifiers
              1..5177
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="HJ03994"
                /clone_lib="pBluescriptII SK plus"
                /dev_stage="adult"
                /sex="male"
                /tissue_type="brain"
              1..2558
                /gene="KIAA0648"
              <1..2558
                /gene="KIAA0648"
              /codon_start=3
              /product="KIAA0648 protein"
              /protein_id="BAA31623.1"
              /db_xref="GI:3327110"
              /translation="LYLVASIDPNNAVKALNEMWKQCNMLRSHVRELLDLHKOPTSEA
              NCSAMFGKLMWIAKNLPDGPKAQDFVKFNQVLGDDEKLRSLQELLELLSPCSCKQADI
              CVREIRKLANPQPTNPLFEMVKFLERIAPIYHDSIEASIALVKLNMKNSIEGTADDE
              EGVSDDPTAIRSGLEKLVLSYFTHPTSFHSAPTESLQCLRMEDDKVAEAPATQIFRN
              THGKIDTLPQIRSTLPIHLQKAKRGTPQAKQAVYCHIAIFTRNEQLVAQIPEFLK
              RSLNADPQLQITPLVSLGHSIMLADPQFASPMKSVVNFIVKDLLMDNRSTGCEKLS
              LKSPDEVSPEVLKQVAKLLVRLWLLGMKNKQNSKANTSLRLLSAMLVSEGLDTQK
              RISKSDSRLRLAAGSAIKALQPCYCHELIPEQFOLCALVINDECYQVROIFPAQL
              HKALVKLLPLEYMAIFALCAKDPVERAHARQCLLNKINISIRREYKQNPMAFERLL
              SLUPEYVTPVPHLLADPDTFSDQVDLQDIKECLWFLEMLVIMTKNNNSHAFMKR
              MAENIKLTDQAQSDPDESKNEKLYITCDVALCVINSKALCNADSPKDPVLPMKHFTQ
              PERDFCNDSYIIEETFRVLLTGKPKPAGVLGAVNKPGLSATGRKPVVSTGTGTGNSI
              NVNSLFGTQVNRSSQSAATGVSENEENPVRIISVTPVKNIIDPVKNKEINSQDA
              TCGNISDSDGKKFTVTAAGAENIQOKTDEKVDDESGPPAPSKPRGRPKSESOGNATK
              NDLNLRPINGKRRAAVGQSPSGLEAGNKAFLQDLAKKAAPAERQIDLQR"
BASE COUNT  1690 a 931 c 1037 g 1519 t
ORIGIN

Query Match          19.88; Score 1041.4; DB 85; Length 5177;
Best Local Similarity 68.08; Pred. No. 1.6e-184;
Matches 1450; Conservative 0; Mismatches 681; Indels 0; Gaps 0;

QY  1492  gcttatattctgtatgccacactggattaaatcgtgtgaagcattgaatgaaatgt 1551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    1  GCTATATTACTATATGCTAGTTTGGATCCAAATGCTGTGAAGACTCTCAACGAATGT 60

QY  1552  ggaatgtcaaaatctgtccgcacatcaagtaagagatttgcttgacttattaagcaac 1611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    61  GGAAGTGTGAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTGGACAAAGCAGC 120

QY  1612  ccaaaacagatgccagtggtcaaggccatattttcaaaagtgatggttattacaagaatt 1671
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    121  CTACATCAGAGGCTAACTGTTTGCCCATGTTTGGAAAACCTGATGACCATAGCAAGAAGATT 180

QY  1672  tactgtacctggttaaggctcaggtatttcataagaataattcacacaggtgttagaagatg 1731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    181  TGCCTGACCCCGGAAGACACAGATTTTGTGAAGAANTTTAACACAGTTCTCGGCCGATG 240

QY  1732  atgagaaataaagaagcagttagaagacttgtagtccaaacatgctctctgcagcagg 1791

```







```
QY 3697 tggagagccctagagcagagaaacaccccgctcacagaaacaggagagaaattaggta 3756
Db 3541 TGGAGAGCCCTAGAGCGAGGAGAAACCGCCGCTCAGAACAGGAGGAGAAATAGGTA 3600
QY 3757 tggatgacttactaagtggtagagaaacagaaacctaaagcagtcagcgaagtcgga 3816
Db 3601 TGGATGACTTACTAGTTGGTACAGGAACAGAAACCTAAAGCGAGTCAGCGAAGTCGGA 3660
QY 3817 aaagagcccatcagcgcttcagaatctgtagaacagcagtcggcctgaggaagaggtctca 3876
Db 3661 AAAGAGGCCATACGGCTTCAGAAATCTGATGAACAGCAGTCGGCTGAGGAAACAGAGGCTCA 3720
QY 3877 aaagaatataattagaaataagagtagaacagaaatagtcgcgcacaaagggtaaaagag 3936
Db 3721 AAGAAGATATATTAGAAATGTAAGATGAACAGAAATAGTCGCCCAAAAGGGTAAAGAG 3780
QY 3937 gccgacccacaaacctcttggctggaggtacaccaaagaagcgaacaaatgaaacctt 3996
Db 3781 GCCGACCACCAAACTCTTGGTGGAGGTACACCAAAAGAGAGCCACCAATGAANAATT 3840
QY 3997 ctaaaaaggaagcaaaaaaatctggacctccagaccagagggaggaagaagaag 4056
Db 3841 CTAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACAGAGGAGGAGGAAGAAAG 3900
QY 4057 aaagacaaagtgaatacaggaacagaagtcacaaagcaaacagaccgagtgcaagga 4116
Db 3901 AAAGCAAAAGTGGAAATACAGGAACAGAAAGTCCAAAGCAACAGCACCGAGTCAAGGA 3960
QY 4117 gagcacagagagagcagaaatccctgaatctagtgcaatcgaatccacacagtcacac 4176
Db 3961 GAGCAGACAGAGAGAGATCTCTGAAATCTAGTGAATGAATGAATCCACAGTCCACAC 4020
QY 4177 cacagaaagcaggaagaaccatcaaaacgcgcacatccaccatcacacacaaacaaaaatg 4236
Db 4021 CACAGAAAGGACGAGGAAGACCATCAAAACGCCATCACCATCACAAACCAAAAAAATG 4080
QY 4237 tgtaagttgtaaatattacattcaa---accaatttcaaatattttgcaaaagtccct 4293
Db 4081 TCCGTGTAGACGCTCCAAACAGCAGCTACTAAGGAAATGATTCAAGTGAAGAAAGTAG 4140
QY 4294 aaattgtaaacatacatattgctgtatttaaatccata-----tatt 4337
Db 4141 ATGTGTTTCAGGTAGCTCTCTGTCGATGATATTCACAGGAAGAAACAGAGGAGGAGG 4200
QY 4338 tagccccattactaggtacgagcgcgaagtgtcaaaaggggaacgcgatgaacaaatg 4397
Db 4201 AAGTTCTACAGTAAATGTACGGCGGAAGTGTAAAGGGGAACGCGATGAACAAATG 4260
QY 4398 taattaataacttctctgtaagcttttgaaaaatcttttttttttttttttttttttt 4457
Db 4261 TAATTATAACTTCTCTGTGGAAGCTTTGGAAAAATC-----TTTTTTTTTTTTTTTTTT 4317
QY 4458 ggtcaagcttgagcgtgaataaaagcctttgatgcacaaaatggagactgctgaagagtgga 4517
Db 4318 GGTCAAGCTTGAGCGTGAATAAGCCTTTGATCCACAAATGGGATGGCTGTGAAGATGGA 4377
QY 4518 cagttggaccttactttggtagccccatacatatttggtaacatgcttttagccatacaca 4577
Db 4378 CAGTTGGACCTTACTTTGGTGACCCCATACGTTTGTGGTCACATGCTTTAGCCATACACA 4437
QY 4578 tggtaacattgactggagcttggaaagtgtgaaagttaagtgcgaatgctatgtagacataa 4637
Db 4438 TGGTAACTTTGACTATGGAGCTTTGTGAAAGTGTAAATGTGCGATGGCTATGTAGACATAA 4497
QY 4638 agaagaacctgtaaatacttttttttttttttttttttttttttttttttttttttttt 4697
Db 4498 AGAAGAACTGTAAATATCTTTTCTTTTCTTTTAAATGTTCTGATTTCTGAAGTGCYT 4557
QY 4698 gtatagcttttctcggcgctttaaactgacagtaacccgactgtttattggtatttga 4757
Db 4558 GTATAGCTTTTATCTCGCGCTTTAAACTGACAGTACCCGACTGTTTATTGATCTATTGA 4617
QY 4758 ttgaaaaagaatttggtagatagatctttaaagcagtaaatctgacagtggtttgtatttga 4817
```

```
Db 4618 TTTGAAAAGAAATTTGTAGGATAGATCTTTAAGCAGTAATCTGCAGTGTGTATTGTGA 4677
QY 4818 ttcttgcaatttactgtgaaaaaaattgttttcaacaattgggtgctatttcttga 4877
Db 4678 TTCTCTGCAATTTTACTGTGAAAAAAATTTGTTTTCACAAATGGTGCTCATTTTCTTGA 4737
QY 4878 tgtcaatttgttgagagtgtaaatggctctctcccttttggtagtatcttaccctagtgtt 4937
Db 4738 TGTCACTATTGTTGGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACCTAGTGTCTT 4797
QY 4938 actcctgggcacaccttaattctcagaggggtgctaaattgtctgcccattacaccagaagat 4997
Db 4798 ACTCCTGGGCACCCCTTAATCTTCAGAGGTGCTAAATTTGTCTGCCATTTACACCAGAAGAT 4857
QY 4998 gccctctgtaggaggaacacacatgcaaatgtgaaatagtcctcctggaagttcttggattact 5057
Db 4858 GCCTCTGATAGGAGGACACACCATGCAAAATGTGAAATAGTCTTGAAGTCTTGGATTAAT 4917
QY 5058 ttacacctcagtagtattgttgcacagaattttctggcctttcatggcaatgaaatttt 5117
Db 4918 TTACACTCAGTATTGATTTGTGCCAGAAATTTCTGCCCTTTTCATGGCAATGAAAAATTTT 4977
QY 5118 aagaagaagatttaagtagtattttaaagagtggttttaaaagagtggttataaaataatgactgaa 5177
Db 4978 AAGAAGAAAGATTTTAAAGTATTTTAAATTTTAAAGAGTGTGTTATATAAAATGTTACTTGAA 5037
QY 5178 ttctttatcccatcttatcatcctcttcagtttttataatctactgtatcaataaaattc 5237
Db 5038 TTCTTTATCCCATTTTATCATCTCTTTCAGTTTTTATTAATCTACTGTATCAATAAAATTC 5097
QY 5238 tgtaatttgaatg 5250
Db 5098 TGTAAATTTGAATG 5110

RESULT 4
HSU50533 1852 bp mRNA PRI 27-NOV-1996
LOCUS Human BRCA2 region, mRNA sequence CG008.
DEFINITION U50533
ACCESSION U50533
VERSION U50533.1 GI:1531605
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1852)
AUTHORS Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,
Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,
Farid,L., Frye,C., Hattier,T., Janekci,T., Jiang,P., Kehrer,R.,
Leblanc,J.F., McArthur-Morrison,J., McSweeney,D., Miki,Y.,
Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,
Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,
Tranchesi,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,
Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,
Tavtigian,S.V. and Simard,J.
Generation of an integrated transcription map of the BRCA2 region
on chromosome 13q12-q13
JOURNAL Genomics 36 (1), 86-99 (1996)
MEDLINE 96411650
REFERENCE 2 (bases 1 to 1852)
AUTHORS Simard,J.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular
Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,
Quebec City, Quebec G1V 4G2, Canada
FEATURES
source
1. .1852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
```

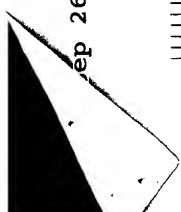












Db	3482	TTTTCACTCTCTGAAAAACCTAAACAAACAAATGTTCTAGAGCTGTTAAACAGGCACATTT	3541
Qy	3481	catcagcaggcaagcaatctcagaccacaatcatcagcaatggaactgtaagcaatgcaa	3540
Db	3542	CATCAGCAGGCAAGCAATCTCAGCCCAATCATCATCAATGGAATCTGTAAGCAATGCAA	3601
Qy	3541	gcagcagctcaaatccaaagctctccctgggaagataaaggggaggttgatgtcttgaaa	3600
Db	3602	GCAGCAGCTCAATCCAAAGCTCTCTGGAAGATAAAGGGAGGCGCTTGATAGTTCTGAAA	3661
Qy	3601	tgatcacagtgaaaaatgaagattacacaatgtcttcacactttgcccgggaaaaaaagtg	3660
Db	3662	TGGATCACAGTGAATAATGAAGATTACACAAATGCTTCACCTTTGCCGGGAAAAAAGTG	3721
Qy	3661	acaagaagacgactctgtatctgttaaggtctgaaatggagaagcctagagcgagaaaa	3720
Db	3722	ACAAGAGAGAGACTCTGATCTTGTAAAGTCTGAATTTGGAGAAGCCTAGAGCGCAGGAAAA	3781
Qy	3721	aaacgcccgtcacagaacagggagagaaattaggtatggatgacttgactaagtgtgtac	3780
Db	3782	AAACGCCCGTCACAGAACAGGAGAGAAATTAGGTATGGATGACTTCACTTAAGTTGGTAC	3841
Qy	3781	aggaaacaaaccttaaggcagtcagcgaagtcgaaagagccatacagcgtctcagaat	3840
Db	3842	AGSAACAGAAACCTTAAAGGCAGTCAGCGAAGTCGGAAGAGAGGCCATACGCGTTCAGAA	3901
Qy	3841	ctgatgaacagcagctggcctgaggaagagaggtccaagaagatatattagaagaatgaag	3900
Db	3902	CTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAGAAGATATATTAGAAAATGAAG	3961
Qy	3901	atgaacagaatagtcgcgcaaaaagggttaaaagagccagccacaaacacctcttggtg	3960
Db	3962	ATGAACAGATAGTCCGCAAAAAGGTTAAAGAGGCGCCAGCCACCAAAACCTCTTGGTG	4021
Qy	3961	gaggtacaccaaaagaagcagcaaatgaaactcttaaaaagggaagcaaaaaaaat	4020
Db	4022	GAGGTACACCAAAAGAAGAGCCACAAATGAATACTCTTAAAAAGGAAGCAAAAAAAT	4081
Qy	4021	ctggacctccagcaccagagaggagggaagaagaagaagcaaatggaataatcaggaac	4080
Db	4082	CTGAGACCTCCAGCACAGAGGAGGAGAGAGAGAGAAAGAAAGTGGAAATACGGAAC	4141
Qy	4081	agaagtcacaaaagcaaacagaccagctgcaaggagagcagcagagagcagaaatctc	4140
Db	4142	AGAAGTCCAAAAGCAACAGCACCGAGTGTCAAGGAGAGCAGCAGAGAGCAGAATCTC	4201
Qy	4141	ctgaactagtgaattgaatccacacagtcacacagccacagaaagagcagaggaagaccat	4200
Db	4202	CTGAATCTAGTGAATTGAATCCACACAGTCCACACACACAGAAAGGAGGAGAGACCAT	4261
Qy	4201	caaaaagcgcatacacatcacacaaaataatgtaagtgttaaatattacatttc	4260
Db	4262	CAAAAAGCGCATACCATCACAAACCAAAAAAATGTCGCTGAGGAGCTCCCAACCAAG	4321
Qy	4261	aa---accatttcaaatatttttgcacaaagtctctaaatttgttaacatacatattgct	4317
Db	4322	CACCTACTAAGGAATGATTCAAGTGAAGAGTAGATGTGTTTCAGGGTAGCTCTCCTG	4381
Qy	4318	gtatttaattccatat-----attagccccattacactaggtacgcg	4361
Db	4382	TCGATGATATTCACAGGAGAAACAGAGAGAGGAGAAATTTCTACAGTAATGTACGGC	4441
Qy	4362	gcgaagtcctaaaaggagcgcatgaacaaatgtaatttaataacttctctgtgaaa	4421
Db	4442	GGGGAAGTGTAAAAGGAACGCGATGACAAATGTAATTAATTAATCTTCTGTGAAA	4501
Qy	4422	gcttggaaaaatcttt	4481
Db	4502	GCCTTGAAAAATCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT	4561
Qy	4482	cccttgatcacaaaatgggactgctgaagatggacagcttgacacttcttggtagcc	4541

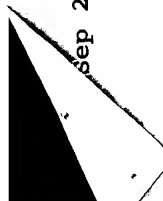
RESULT 3  
AB023196 LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.  
AB023196 ACCESSION  
AB023196.1 GI:4589601  
Homo sapiens adult male brain cDNA to mRNA, clone\_lib:paluescriptII  
SK plus clone:hj07056.  
Homo sapiens ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (sites)  
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,  
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
Prediction of the coding sequences of unidentified human genes.  
XIII. The complete sequences of 100 new cDNA clones from brain  
which code for large proteins in vitro  
DNA Res. 6 (1), 63-70 (1999)  
99246063  
2 (bases 1 to 5110)  
Ohara,O., Nagase,T. and Kikuno,R.



Qy 1261 acaaacgagtggaagtacgcaagaagccatgatgggacttgcacaaatttataagaat 1320  
Db 1322 ACAACGATGGAGAGTACGCAAGAAGCCATGATGGGACTTGCACAAATTTATAAGAAAT 1381  
Qy 1321 atgctttacagtcagcagctgggaaagatgctgcaaaacagatagcatggaaccaagaca 1380  
Db 1382 ATGCTTTACAGTCAGCAGCTGGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAGAGCA 1441  
Qy 1381 aattgtacatatattatcaaaaaagtagtattgatgatcgactacttggtgaacggatct 1440  
Db 1442 AATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTGAAACGGATCT 1501  
Qy 1441 ttgctcaatacatggttctcccaatttagaaactacagaacggatgaatgcttatatt 1500  
Db 1502 TTGCTCAATACATGTTTCCCTCACAAATTTAGAAACTACAGAACCGATGAATGCTTATATT 1561  
Qy 1501 acttgatgcacactggatttaaatgctggaagcatggaatgaaatggaatgtc 1560  
Db 1562 ACTTGATGCGCACACTGGATTTAAATGCTCTGAAAGCATTGAATGAATGTGGAATGTC 1621  
Qy 1561 aaatctgctccgacatcaagtaaaagatttgctgaacttgattgaacaccccaaaacag 1620  
Db 1622 AAAATCTGCTCCGACATCAAGTAAAGGATTGCTTGACTTGAATTAAGCAACCCCAAAACAG 1681  
Qy 1621 atgccagtgccaagccatattttcaaaagtgtggttattcaaaagaaatttacctgac 1680  
Db 1682 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATC 1741  
Qy 1681 ctggtgaaggctcaggatttcataagaataatccacacagggtggttagaagatgtagagaaa 1740  
Db 1742 CTGCTAAGGCTCAGGATTTTATGAAGAAATTCACACAGGTGTTAGAAAGATGATGAGAAAA 1801  
Qy 1741 taagaagcagttagaagtagtctttagtcaacatgctcctcaacagcgtgaaggtt 1800  
Db 1802 TAAGAAAGCAGTTAGAAAGTACTTGTAGTCCAAATGCTCTGCAACAGCTGGAAGTT 1861  
Qy 1801 gtgtgctgtaataactaaagaagtgtggcaaccccaacagcctcacaattcctttctg 1860  
Db 1862 GTGTGCGTGAATTAACTAAGAAGTTGGCAACCCCAACAGCCTTACAAATCTTTCCTGG 1921  
Qy 1861 aaatgataagttctcttgagagtagcacctgtgacatagatcaccgaatctatca 1920  
Db 1922 AAATGATCAAGTTTCTCTGGAGAGTAGCACCTGTGCACATAGATACCGAATCTATCA 1981  
Qy 1921 gtgctcttatcaacaagtgcaaaaacaaatagatggaacagcagatgatgaagtgag 1980  
Db 1982 GTGCTCTTTTAAACAGTGAAACAAATCAATAGATGGAAACAGCAGATGATGAAGATCAGG 2041  
Qy 1981 gtgttccaactgataaagccatcgagcagggtcttggaactgcttaaggtagctctcat 2040  
Db 2042 GTGTTCCAACGTGATCAAGCCATCAGCAGGCTTGAACCTGTTAAGGTACTCTCATTTA 2101  
Qy 2041 cacatcccatctcatctctctgaacatctgaaatctgaatcattactggtgctgtgaaa 2100  
Db 2102 CACATCCCATCTCATTTCTCTGCTGAAACATTTGAATCATATTACTGGCTTGTCTGAAA 2161  
Qy 2101 ttgatgtgaaaaagtagcagaagctgcactacaaatttcaaaaacacaggaagcaaaa 2160  
Db 2162 TGGATGATGAAAAAGTAGCAGAAGCTGCATACAAATTTTCAAAACACAGGAAGCAAAA 2221  
Qy 2161 ttgaagaggattttccacatcagatcagcctgtcttccctgttttacatcaaaaacta 2220  
Db 2222 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCCTGTTTATACATCACAAATCTA 2281  
Qy 2221 aaaaaggacccccgcgtcaagcctaataatgccattcattgtatccatcgatattttcta 2280  
Db 2282 AAAAAGGACCCCGCTCAGGCCAAATATGCCATTCATTTGATCCATCGGATATTTTCTA 2341  
Qy 2281 gtaagagaccagtttgacagatatattgagcctctgcataaagagcctagatccaaagca 2340  
Db 2342 GTAAAGAGACCAGTTTGACACAGATATTTGAGCCCTCTGCATTAAGAGCCTAGATCAAGCA 2401

Qy 2341 acctggaacatctcataacaccatttggttactatgtggcatattgctctctctgcaactg 2400  
Db 2402 ACCTGGAAACATCTCATACACCATTGGTTACTATTGGTCATATTGCTCTCCTTCACCTG 2461  
Qy 2401 atcaatttgctgctccttggaattcttgggtagctactttcatgtgaaagatctctca 2460  
Db 2462 ATCAATTTGCTGCTCTTGAATCTTTGGTAGCTACTTTTCATTTGTGAAAGATCTTCTCA 2521  
Qy 2461 tgaatgatcggtctccagggaagaaacaaactaaactttgggttccagatgaagaagtat 2520  
Db 2522 TGAATGATCGGCTTCCAGGAAAAACAACTAAACTTTGGTTCCAGATGAAGAAGTAT 2581  
Qy 2521 ctcttgagacaatggtcgaattcaggctattaaatgatggttcgatggctactgtgaa 2580  
Db 2582 CTCCTGAGACAAATGCTCAAAATTCAGGCTATTAAATGATGGTTGATGGCTACTTTGGAA 2641  
Qy 2581 tgaataataatcacagttaaatcaggaaactctaccttaagattgttaacaacaaatattgc 2640  
Db 2642 TGAATAATTAATCACAGTAAATCAGGAACCTTCTACCTTTAAGATTGCTAACACAAATATTGC 2701  
Qy 2641 atagtgtaggagacttgacagacaggggaaaaatttagtaaaccagatatgtcaactctga 2700  
Db 2702 ATAGTGATGGAGACTTGACAGAACAGGGGAAAAATTTAGTAAACACAGATATGTCACGCTCA 2761  
Qy 2701 gacttgctgctgggagtgctattgtgaagctggcacaagaacccctgttaccatgaaatca 2760  
Db 2762 GACTTGCTGCTGGGAGTGTCTATTGTGAAGCTGGCACAGAACCCTGTTACCAATGAATCA 2821  
Qy 2761 tcacattagaacaatatcaagctatgtgcattagctatcaacgatgaatgctatcaagtaa 2820  
Db 2822 TCACATTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAA 2881  
Qy 2821 gacaaagtgttgcacagaaacttcacaaaggccttcccggtttacggcttccacttgagt 2880  
Db 2882 GACAAAGTGTGTGCCAGAACTTCACAAAGGCTTTCCTCGTTTACGGCTTCACCTGAGT 2941  
Qy 2881 atatggcaactgtgcccttcttgcaaaagatccctgtaaaagagagaagatcatgctca 2940  
Db 2942 ATATGCAATCTGTGCCCTTTGTGCAAAAGATCTCTGTAAGGAGAGAAAGAGCTCATGCTA 3001  
Qy 2941 ggcaatgttggtagaaaaataaaatgtaagggcgagtatcttgaaagcagcagatgaagctg 3000  
Db 3002 GGCAATGTTTGGTGAATAATATAAATGTAAAGCGGAGTATCTGAAAGCAGCATGCAAGCTG 3061  
Qy 3001 ttatgtaaaaaattatgtctcttaccagagtattgttccatatcaaatccacttt 3060  
Db 3062 TTAGTGAATAATTTATTGTCTCTTCTACAGAGTATGTTGTTCCATATACAAATTCACCTTT 3121  
Qy 3061 tggcacatgacccagattatgtcaaaagtacaggatatgtgaacaaacttaagaatggttaag 3120  
Db 3122 TGGCACAATGACCCAGATTTATGTCAAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAG 3181  
Qy 3121 aatgtcttgggttcttggaataatttaatggctaaaaatgaaataaagcagtcacgctt 3180  
Db 3182 AATGTCTTTTGGTTGTCTGGAATATTAAATGGCTAAAAATGAAATGAAATGACGTCAGCTT 3241  
Qy 3181 ttatcagaagatggttagaaaaatttaacaacaaagaatgcccaaggagaccagatgatg 3240  
Db 3242 TTATCAGAAAGATGGTAGAAAAATTTAAACAAACAAAGATGCCAAGGACCATGATGATG 3301  
Qy 3241 caaaaaatgaatgaataaactgtacactgtgtgtgatgttgccatgaatatcatcatgcaa 3300  
Db 3302 CAAAAATGAATGAAAAACTGTACATGTGTGTGATGTGGCATGAATATCATCATGTGTCAA 3361  
Qy 3301 agagtactacacagtttggaatctcctaagaacccggtactaccagctcggtttcttca 3360  
Db 3362 AGAGTACTACATACAGTTTGGAAATCTCCTAAAGACCCGGTACTACCAGCTCGTGTCTTCA 3421  
Qy 3361 ctcaactgacaagaatttcagtaaacaccacaaatattatctgctccctgaaatgaatcat 3420  
Db 3422 CTCAACTGACAAGAAATTTAGTAAACCAACCAAAATTTATCTGCCTCTCTGAAATGAATCAT 3481  
Qy 3421 ttttactcttggaacacccctaaacaaacaaatgttcttaggagctgtttaaacaagccacttt 3480





misc\_feature /number=33  
exon 4255..4307  
misc\_feature /note="matches EST T64867 from clone 22264"  
exon 4299..4434  
misc\_feature /number=34  
exon join(4398..4533,4518..4660,4659..4675,4674..4693)  
misc\_feature /note="matches EST H44781 from clone 188032"  
exon 4435..7444  
misc\_feature /number=35  
misc\_feature 4537..5022  
misc\_feature /note="matches EST A1681422 from clone IMAGE:2272632"  
misc\_feature 4537..4892  
misc\_feature /note="matches EST AA911125 from clone IMAGE:1416862"  
misc\_feature 4538..4969  
misc\_feature /note="matches EST A1275866 from clone IMAGE:1877482"  
misc\_feature 4538..4947  
misc\_feature /note="matches EST AA987361 from clone IMAGE:1603350"  
misc\_feature 4538..4943  
misc\_feature /note="matches EST A1299646 from clone IMAGE:1898010"  
misc\_feature 4562..5003  
misc\_feature /note="matches EST AA447433 from clone 784555"  
misc\_feature 4571..4854  
misc\_feature /note="matches EST H93424 from clone 220700"  
misc\_feature join(4573..4649,4650..4815)  
misc\_feature /note="matches EST W20070 from clone 306049"  
misc\_feature complement(4629..4937)  
misc\_feature /note="matches EST AW089775 from clone IMAGE:2593146"  
misc\_feature 4745..4994  
misc\_feature /note="matches EST AA563884 from clone IMAGE:1013990"  
misc\_feature /note="matches EST AA747568 from clone IMAGE:1269696"  
misc\_feature complement(4803..5332)  
misc\_feature /note="matches EST A1911784 from clone IMAGE:2329873"  
misc\_feature complement(4828..5296)  
misc\_feature /note="matches EST AA563884 from clone IMAGE:1013990"  
misc\_feature 4861..5243  
misc\_feature /note="matches EST D20082 from clone pm1283."  
misc\_feature complement(join(4872..4921,5281..5332))  
misc\_feature /note="matches EST AW212954 from clone IMAGE:2647801"  
misc\_feature join(4872..4921,5281..5332)  
misc\_feature /note="matches EST AA177347 from clone 621810"  
misc\_feature complement(4872..4921)  
misc\_feature /note="matches EST AW048929 from clone  
UI-M-BH1-amo-c-09-0-UI"  
misc\_feature 4872..4921  
misc\_feature /note="matches EST AA154932 from clone 599194"  
misc\_feature complement(4876..5333)  
misc\_feature /note="matches EST A1016896 from clone IMAGE:1627886"  
misc\_feature complement(join(4922..4975,5281..5332))  
misc\_feature /note="matches EST A1012649 from clone RPLBA02"  
misc\_feature complement(4922..4975)  
misc\_feature /note="matches EST AA899586 from clone  
UI-R-E0-cy-e-02-0-UI; matches EST A1172172 from clone  
RMUBV82"  
misc\_feature complement(4931..5332)  
misc\_feature /note="matches EST AA625961 from clone 745478"  
misc\_feature complement(4951..5332)  
misc\_feature /note="matches EST A1140790 from clone IMAGE:1566222"  
misc\_feature complement(5021..5309)  
misc\_feature /note="matches EST H45693 from clone 188292"

Query Match 96.5%; Score 5084.4; DB 92; Length 7444;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 5172; Conservative 0; Mismatches 81; Indels 19; Gaps 2;

Qy 1 ccgagagcccgagtgagcgagtagcgagtcggaaccgaggggtagaaatttt 60  
Db 62 CCGGAGAGCCCGGAGTGACGGAGTAGCGAGTCGGCAACCCGAGGGGTAGAAATATT 121  
Qy 61 ctgtcatggtcattcaagaactaggaccatgatggaataataacatacccgctgggg 120  
Db 122 CTGTCATGCTCATCAAGACTAGGACCAATGATGAAAAAATACATATCCGCTGGG 181  
Qy 121 tcaaggaaatcagataaaatatctaaagagagagatggtagacgattaaatggtt 180  
|||||

Db 182 TCAGGAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTG 241  
Qy 181 tgaaccttttattgatatggaccagactctgaagaagaagaagagctttatttaacc 240  
|||||  
Db 242 TGAAGAACCTTTATGGATATGACCGAGACTCTGAAGAAGAAAGAGAGCTTTATTAAACC 301  
Qy 241 tagctttacatcttgcttcagatcttttttccagacatccctgggttaaagatgttcgcttac 300  
|||||  
Db 302 TAGCTTTATACATCTTGCTTCAGATTTTCTTCAAGCATCTCTGATANAAGATGCTCGCTTAC 361  
Qy 301 tggtagcctgctgcttgcctgatattttcaggattatttgcctcctgaagcctcctaacat 360  
|||||  
Db 362 TGGTAGCTGCTGCTGCTGCTGATATTTTTCAGATTTTATGCTCTCTCAAGCTCCTTACACAT 421  
Qy 361 cccctgataaaactaagatataattttattgattataaagaagacagctgaagggcttagagg 420  
|||||  
Db 422 CCCCTGATAAACTAAGGATATATTATTTATTAACAAAGACAGTGTGAAGGGGCTAGAGG 481  
Qy 421 atacaagagcccaataatcaataggtatttttatttacttgagacattgtcgtgggtca 480  
|||||  
Db 482 ATACAAGAGCCCCACATTCATAGSTATTATTTTACTTGAGAACATTTGCTTGGGTCA 541  
Qy 481 agtcataacatatagtcttggagttagaagatagcaatgaaattttcaccagctatata 540  
|||||  
Db 542 AGTCATATAACATATGCTTTGAGTTAGAAAGATAGCAATGAAATTTTCACCCAGCTATACA 601  
Qy 541 gaaccttatttcagttataaaacatggcccaatcagaagaagtcctatatgcacatggtag 600  
|||||  
Db 602 GAACCTTATTTTTCAGTTTAAACATATGGCCCAATCAGAAATGCCATATGACATGGTAG 661  
Qy 601 acctatgagctctatttttgaagggtgatacagctgctcagagagcttttgggatacgg 660  
|||||  
Db 662 ACCTTATGAGCTCTATTATTTTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGG 721  
Qy 661 ttttagtaaatcggtaacctgctcataaataatttaacaagaagcatatgatttggcaa 720  
|||||  
Db 722 TTTTAGTAAATCTGGTACCTGCTCATAGAAATTTTAAACAGCAAGCATATGATTGGCAA 781  
Qy 721 aggtcttactgaagagagacagctcaagctattgagccatataataccacttttttaac 780  
|||||  
Db 782 AGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACMAATTTTTTAATC 841  
Qy 781 aggtcttgatgcttgggaaaaacatctatcagcagattgtcagagcagctgttgaactaa 840  
|||||  
Db 842 AGGTTCTGATGCTTGGGAAAAACATCTATCAGCGATTGTGTCAGAGCATGCTTTGACTTAA 901  
Qy 841 ttttgagctctacaattatgatagtcattgtgctgctctctgttttaccocagcttgaat 900  
|||||  
Db 902 TTTTGGAGCTCTACAATATTGATAGTCAATTTGCTGCTCTCTGTTTTACCCAGCTTGAAT 961  
Qy 901 ttaataaagagcaatgataatgagagcgctcacaagttgtttaaactactggcaaaaa 960  
|||||  
Db 962 TTAATTAAGAGCAATGATTAATGAGGAGCGCTTACAAAGTTGTTAAACTACTGGCAAAAA 1021  
Qy 961 ttttgggggaaaagattcagaattgcttctcaaaacaaagccactttggcagtgact 1020  
|||||  
Db 1022 TGTTTGGGGCAAGGATTTCAGAAATGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACT 1081  
Qy 1021 tgggcaggttttaataatgatcccatgtaccatcccgctgggaatgtgtgaaaatttctagcc 1080  
|||||  
Db 1082 TGGCGCAGGTTTAATCATATCCATGTACCAATCCGCTGGAATGTGTAATTTTCTAGCC 1141  
Qy 1081 attgtcctatgaacacatcctgatttagcaaaagacttaacagagctattctaaagtggagt 1140  
|||||  
Db 1142 ATTGTCTCATGAACCATCTCTGATTATTAGCAAAAGACTTAAACAGAGTATCTTAAAGTAGAGT 1201  
Qy 1141 cacatgacctgaggaagcctattagacatgatttattgtgtcgaatagctacagctgcta 1200  
|||||  
Db 1202 GACATGACCTCTGAGGAAGCTATTAGACATGATGTTATTGTGTCATATAGTAGTGTGTA 1261  
Qy 1201 aaaaagatatctcttggtcgaatgatcacttaacttaattttgtgagagagaaacattag 1260  
|||||  
Db 1262 AAAAGGATATCTCTGCTCAATGATGATCACTTAAATTTTGTGAGAGAGAGAAACATTAG 1321



exon exon exon



4141	CTGAATCTAGTGC	AAATTGAATCCACAGCTCCACACCACAGAAAGCAGGACGAGACCAAT	4200
Db			
	4201	caaaaagccatcaccatcaccaacccaaaaaaatgtgaagtgtgtaaatattacatttc	4260
Qy			
Db	4201	CAAAAAGCCATCACCATCAACACAAAAAAATGTGAAGTGTGTAATATATAGCATTTTC	4260
	4261	aaaccaattccaatattatttggaaaaagttctaaatttggtaaaacatacatattctctga	4320
Qy			
Db	4261	AAACCAAATTTCAAATTAATTTGCAAAAGTTCTTAATATTTGTAACCATACATATTGCTGTA	4320
	4321	tttaaatcccatattttagcccatcacactaggtacgcgcggaagtgtctaaaaggga	4380
Qy			
Db	4321	TTTAAATTCATATATTAGCCCATTACACTAGGTACGGCGCGGAAGTCTAAAAAGGA	4380
	4381	acggcgatgaacaaatgtaataaactttctctgtgaaagcttggaaaaatcttttt	4440
Qy			
Db	4381	ACGGCGATGAACAAATGTAAATAACTTTCTCTGTGAAAGCTTTGGAAAAATCTTTT	4440
	4441	tttttttttttttttggctcaagcttgagctgaaataagacctttgatgcacaaatgg	4500
Qy			
Db	4441	TTTTTTTTTTTTTTTTTGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGG	4500
	4501	gactgctgaagatggaacagttgacaccttactttggtgacccacatacatttgggtcaca	4560
Qy			
Db	4501	GACTGCTGAAGAGTGGACAGTTGGACCTTACTTTGGTGACCCCATACATTTGGTGACA	4560
	4561	tgctttaaccatacacatggtlaaacatgactatgagctctgtgaaagtgtaatgtgcga	4620
Qy			
Db	4561	TGCTTTAGCCATACACATGGTAACATTTGACTATGGAGCTTTGTGAAGTGAATGTGCGA	4620
	4621	tgctatgtagacataaagaagaacactgtaaatatctttttcttttttttaagtcttc	4680
Qy			
Db	4621	TGGCTATGTAGACATAAAGAAGAAACTTGTAAATCTATCTTTTCTTTTTTAACTGTTTC	4680
	4681	tgatttctgaagtgtctgtatagctttttatctcgcgctttaaactgacagcaccgactg	4740
Qy			
Db	4681	TGAATTTCTGAAGTGTGTATAGCTTTTATCTCGCGCTTTAACTGACAGTACCCGACTG	4740
	4741	tttattggatctattgatttgaaaagaaatttgtagtagatgtttaagcagtaattctgt	4800
Qy			
Db	4741	TTTATTGGAUCTATTGATTTGAAAGAAATTTGTTAGGATAGATCTTAAGCAGTAATCTGT	4800
	4801	cagtggttattgtatttctcgaatttacttgtaaaaaaatttgttttcaacaat	4860
Qy			
Db	4801	CAGTGTGTTGATTTGTATTTCTCGAAATTTACTGTGAAAAAAATTTGTTTCAACAAT	4860
	4861	tggtgcatttcttgatgcactattgttggagagttaaatggctcttccttcccttctg	4920
Qy			
Db	4861	TGGTGTCAITTTCTTGATGTCATATTGTTGGAGAGTTAAATGGTCTCTCCCTTTGTG	4920
	4921	tattctacctagtggttactctcctggcacccttaactctcagagtgctcaaatgtctgc	4980
Qy			
Db	4921	TATCTTACCTTAGTGTATTCTCTCGGCACCCCTTAATCTTCAGAGGTGCTAAATTTCTCG	4980
	4981	cattacaccagaagatgcctctgtataggaggaaccaatgcataattgtgaaatagtcct	5040
Qy			
Db	4981	CATTACACCAGAAGATGCCCTCTGTAGGAGGACAAACCATGCAAAATGCTGAAATAGTCCT	5040
	5041	gaagtcttggaatactttacacctcagtatgtattgtcccagaatttctgccttcc	5100
Qy			
Db	5041	GAAGTCTTCGGATTACTTTACACCTCAGTATTGATTTGTGCCAGAAATTTCTCGCCCTTC	5100
	5101	atgccaatgaaattttaagaagaagatttaagattatttaatttaaagagtgtgta	5160
Qy			
Db	5101	ATGCCAATGAAATTTTAAAGAAAGAAATTTAAAGTATTTTAAATTTTAAAGAGTGTGTA	5160
	5161	taaaataatgtactgaaatttcttaccattttatccacttttatactcattttataactta	5220
Qy			
Db	5161	TAAAAATATGTACTGAAATCTTTATGCCATTTTATCATCCTTTTCAGTTTATTATTAATCTA	5220
	5221	ctgtatcaataaaattctgtaatttgaatgagt	5253
Qy			
Db	5221	CTGTATCAATAAATTCGTAAATTTGAATGAGT	5253



QY 901 ttaattaaagagcaatgataatgaggagcgccctacaaagtgtttaaaactactggycaaaa 960  
Db 901 TTAATTTAAAGACGAATGATAATGAGGAGCGCTCAAGTTGTTAAACTACTTGGCGAAAA 960  
QY 961 tgtttggggcaagattcaggaattggctctcaaaaacagccacttggcagtgctact 1020  
Db 961 TGTTTGGGCAAGGATTCAGAAATGGCTTCTCAAAACAGCCACTTTGGCAGTGTCTACT 1020  
QY 1021 tgggcaaggtttaatgatatccatgctaccaatccgcctggaatgtgtgaaatttgcagcc 1080  
Db 1021 TGGGCAAGGTTTAATGATATCCATGTACCAATCCGCTGGAATGTGTGAATTTGTCTACCC 1080  
QY 1081 attgtctcatgaaccatctgatttagcaaaagacttaacagagtatatttaaaagtgaagt 1140  
Db 1081 ATTGTCTCATGAACCATCTGTATTTAGCAAAAGACTTAAACAGAGTATCTTAAAGTGAGGT 1140  
QY 1141 cacatgacctgaggaagctattagacatgatgttattgtgtcaatgattcacagctgcta 1200  
Db 1141 CACATGACCTGAGGAAGCTATTAGACATGATGTTATTGTGCTAATAGTTACAGCTGCTA 1200  
QY 1201 aaaagatatctcttgctaatgatcatcacttaacttaatttbtgtgagagagagaacattag 1260  
Db 1201 AAAGGATATTCTTCTGTGCTAATGATCATCTTACTTAATTTTGTGGAGAGAGAACATTAG 1260  
QY 1261 acaaacgatggagagatcacgaagaagccatgatgggacttgcacaaattttataagaagt 1320  
Db 1261 ACAACGATGGAGAGTAGCGAAAGNAGCCATGATGGGACTTGGCCCAATTTTAAAGAAAT 1320  
QY 1321 atgctttacagtcagcagctggaagaagatgctgcaaacagatagcatggatcgaacagaca 1380  
Db 1321 ATGCTTTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAGCATGGATCAAGACA 1380  
QY 1381 aattgctacatatattatacaaatagattgatgatcgactcttgttgaacggatct 1440  
Db 1381 AATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTGTGAACGGATCT 1440  
QY 1441 ttgctcaatacatgttctcctcaaatattagaacctcagaaactcagagagataaatgttatatt 1500  
Db 1441 TTGCTCAATACATGTTCTCTCACATTTAGNAACTACAGAACGGATGAATGCTTTATATT 1500  
QY 1501 acttgatgcacactggatttaaatgctgtgaagcattggaatggaatgtggaatgtc 1560  
Db 1501 ACTTGTATGCCACTCGATTTAAATGCTGTGAAGCATTTGAATGAAATGTGGAATGTC 1560  
QY 1561 aaactctgctccagacatcaagtaaaagatttgccttaacttgatgaagcaacccaaacag 1620  
Db 1561 AAATCTGCTCCGACATCAAGTAAGGATTTGCTTGACTTTGATTAAGCAACCCCAAAACAG 1620  
QY 1621 atgccagtgtaaggccatattttcaaaagtgtggttattacaagaaattttacctgac 1680  
Db 1621 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAGTTTACCTGATC 1680  
QY 1681 ctggtaaggtcaggtattcatagaagaaattcacacaggtgttagaagatgatgagaaaa 1740  
Db 1681 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTCACAGGTTGTAGAAAGATGATGAGAAAA 1740  
QY 1741 taagaagcaggttagaagtacttgttagtccaacatgctcctcgaagcaggtgtgaaggtt 1800  
Db 1741 TAAGAAAGCAGTTAGAAGTACTTGTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTT 1800  
QY 1801 gttggtgaaataactaagaagttgggcaaccccccaacagactcaaaatcttctctgg 1860  
Db 1801 GTGTGGGTGAATTAACATAAGAGTTGGGAACCCCAACAGCCTACAAATCTCTTCTCG 1860  
QY 1861 aaatgatcaagttctcttggagagatagcactgtgacatagataccgaatctatca 1920  
Db 1861 AAATGATCAAGTTTCTTCTGGAGAGATAGCACCCTGTGCACATAGATACCGAATCTATCA 1920  
QY 1921 gtgctcttataaacaagtgaacaaatcaatagatggaacacagatgatgagatgag 1980  
Db 1921 GTGCTCTTATTAACAAGTGAACAAATCAATAGATGGAAACACAGATGATGAAGATGAGG 1980

QY 1981 gtgttccaactgatcaagccatcagagcaggtcttgaactgttaaggtactctcattta 2040  
Db 1981 GTGTTCCAACGTGATCAAGCCATCAGAGAGGTCTTGAACACTCTTAAGGTACTCTCATTTA 2040  
QY 2041 cacatcccatctcatcttccattctgtctgaacatttgaatcattactgtgtctgtgaaaa 2100  
Db 2041 CACATCCCATCTCATCTCATTTCTGCTGAAACATTTGAATCATTTACTGGCCTGTCTGAAA 2100  
QY 2101 tggatgatgaaaaagtagcagaagctgcactacaaaaattttcaaaaaacacaggaagcaaaa 2160  
Db 2101 TGGATGATGAAAAAGTAGCAGAAGCTGCACATACAAATTTTCAAAAAACACAGGAAGCAAAA 2160  
QY 2161 ttgaagaggattttccacacatcagatcagccttgcctctctgttttacatcacaaatcta 2220  
Db 2161 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCTGTTTACATCACAAATCTA 2220  
QY 2221 aaaaagagaccccccgctcaagcacaatatccattctatgataatcgcgatatcttcta 2280  
Db 2221 AAAAAGGACCCCCCGCTCAAGCCCAAAATATGCCATTTCATTGTATCCATGCCATATTTTCTA 2280  
QY 2281 gtaaaagagacccagtttgcacagatattttagcctctgcataaagagccttagatccaaagca 2340  
Db 2281 GTAAAGAGGCCAGTTTGGCACAGATATTGAGCCTCTGCAATAAGAGCCTAGATCCAAAGCA 2340  
QY 2341 acctggaacatctcatcaaacaccattgttactattgtgtcatattgtctctctctgcacctg 2400  
Db 2341 ACCTGGAACATCTCATAAACACCATTTGTTACTATTGTCATATTGCTCTCTCTTGCACCTG 2400  
QY 2401 atcaatttgcctctcttggaaatcttgggtagctactcttcttcttgaagatctctctca 2460  
Db 2401 ATCAATTTGCTGCTCTCTTTGAAATCTTTGGTAGCTACTTTTCATTGTGAAGATCTTCTCA 2460  
QY 2461 tgaatgatcggtctccagggaaaaagacaactaaaccttgggttccagatgaagaagtat 2520  
Db 2461 TGAATGATCGGCTTCCAGGGAAAAACACAACTAAACTTTGGGTTCAGAGATGAAGAAAGTAT 2520  
QY 2521 ctctcgagacaaatgggtcaaaattcagggttatlaaaatgatgtgtcgtacttctggaa 2580  
Db 2521 CTCCTGAGACAAATGGTCAAAATTCAGGCTATTAAATATGATGTGTCGATGGCTACTTGGAA 2580  
QY 2581 tgaataaataacacagtaaatcaggaactcttaccttaagattgctcaacaaatattgc 2640  
Db 2581 TGAATAATTAATCACAGTAATCAGGAACCTTACCTTAAAGATTGCTTAAACAAATATTTCG 2640  
QY 2641 atagtgtgagacttgacagacaggggaaaaatttagtaaacacagatatgtcacgctcga 2700  
Db 2641 ATAGTGATGGAGACTTGACAGAACAGGGGAAAAATTAGTAAACACAGATATGTCAAGTCTGA 2700  
QY 2701 gacttgcctggaggtgctatttgaagctggcacaagaacccctgttaccatgaaatca 2760  
Db 2701 GACTTGTCTCTGGGAGTGTCTATTGTTGAAGCTGGCACAGAACCCCTGTTACCATGAATCA 2760  
QY 2761 tcacattagaacaataatcagctatgtgcatatagctatcaacgatgaatgctatcaagtaa 2820  
Db 2761 TCACATTAGAACAATATCAGCTATGTGCTATTAGCTATTCAACGATGANTGCTATCAAGTAA 2820  
QY 2821 gacaagttgttgcacagaacttcaaaaggccttcccggtttacgggttccacttggat 2880  
Db 2821 GACAAGTGTGTTGCCACAGAACTTCAAAAGGCTTTTCCCGTTTACGGCTTCCACTTGAGT 2880  
QY 2881 atatggcaatctgtgccccttgtgcaaaagatcctgtaaaaagagagaagactcatgcta 2940  
Db 2881 ATATGGCAATCTGTGCCCTTTGTGCAAAAGATPCTGTAAAGAGAGAGAGACTCATGCTA 2940  
QY 2941 ggcaatgttgggtgaaaaataataatgtaagcgaggagtatctgaaagcagcagtcagctg 3000  
Db 2941 GGCAATGTTGGTGAATAATAATAATGTAAAGCGGGAGTATCTGAAGCAGCATGCAGCTG 3000  
QY 3001 ttagtgaataattattgtctcttctaccagaggtattgtgttccatatacaatctacattt 3060  
Db 3001 TTAGTGAATAATTATTGTCTCTTACCAGAGTATGTTGTTCCCATATACAAATTCACCTTT 3060  
QY 3061 tggcacatgacccagatttatgtcaaaagtacaggatatattgaacaactttaaagatgttaaag 3120



exon  
4130..4354  
/gene="AS3"  
/number=33

exon  
4355..4810  
/gene="AS3"  
/number=34

BASE COUNT 1755 a 944 c 1074 g 1480 t

ORIGIN

Query Match 99.1%; Score 5221; DB 97; Length 5253;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 5233; Conservative 0; Mismatches 20; Indels 0; Gaps

QY 1 ccgagagcccgagtgagcggagtagcgagtgcggcaaccggagggtgagaaatattt 60  
|||||  
Db 1 CCGGAGACCCGGAGTGAGCGGAGTAGCGGAGTAGCGCACGCCGAGGGGTAGAAATATT 60  
|||||

QY 61 ctgtcatgcttcattccaagaactaggaccaaatgatggaaaaattacatatccgcctggg 120  
|||||  
Db 61 CTGTCATGCTCATTTCAAGACTTAGGACC AATGATGSA AAAATTACATATCGCCTGGGG 120  
|||||

QY 121 tcaaggaaatcacagataaaatactcaacgagagatggtgagacqattaaagatggttg 180  
|||||  
Db 121 TCAAGGAATAATCAGATAAAATATCTTAAAGAGAGATGGTGAGCAGATTAAAGATGGTTG 180  
|||||

QY 181 tgaaaacttttatgatgatgaccaggacctctgaagaaagaaaggagctttatttaaac 240  
|||||  
Db 181 TGAAAACITTTATGATATGACCCAGGACTCTGAAGAAGAAAAAGAGAGCTTTATTTAAACC 240  
|||||

QY 241 tagctttaacatctgtctcagatatttttctcaagcatcctgttaaagatgttcgcttac 300  
|||||  
Db 241 TAGCTTTACATCTTGCTTCAGATTTTTTTCTCAAGCATCTGTATAAAGATGTTTCGCCTTAC 300  
|||||

QY 301 tgttagcctgtgccttgtctgatatttccagatttatgtcctcgaagctccttcacacat 360  
|||||  
Db 301 TGGTAGCCCTGCTGCCTGTGATATTTTCAGGATTTATGTCTCTGAAGCTCCTTTACACAT 360  
|||||

QY 361 ccctgatataaactaaagatatatttatgtttataacaagacagttgaagggtctagagg 420  
|||||  
Db 361 CCCTGATAAACCTAAGGATATATTATGTTTATAACAAGACAGTTCTGAAGGGCTTAGAGG 420  
|||||

QY 421 atacaaggccccacaattcfaataggattttttatttacttgagaacattgcttgggtca 480  
|||||  
Db 421 ATACAAGAGCCCCACAGTTCATAGTAATTTTATTTGCTTGAGACATTTGCTTGGTGCA 480  
|||||

QY 481 agtcataacatatgtctttgagtgtagaagtagaatgaaatttcccacagctataca 540  
|||||  
Db 481 AGTCATATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCCCACAGCTATACA 540  
|||||

QY 541 gaacctattttcagttataaacaatgcccaatcagaagtcctatgcacatggttag 600  
|||||  
Db 541 GAACCTTATTTTCAGTTATAACAATATGGCCACATTCAGAAAGTCCATATGCACATGGTAG 600  
|||||

QY 601 accttagagctctattttgtgaagtgatcacagttgtccaggagcttttggatcagg 660  
|||||  
Db 601 ACCTTAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGG 660  
|||||

QY 661 tttagtaaatctggtacctgtctcataagaattataacaagacagcatagtttggcaa 720  
|||||  
Db 661 TTTTAGTAAATCTGGTACCTGCTCAATAAGAAPTTTAAACAAGCAAGCATATGATTTGGCAA 720  
|||||

QY 721 aggccttactgaagagagcagctcgaagctatttgaccctatatlaccacacttttttaac 780  
|||||  
Db 721 AGGCTTTTACTGAAGAGCAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTTAATC 780  
|||||

QY 781 aggttctgatgcttgggaaaaacatctatccagcagatttgcagagcatagtcttgaactaa 840  
|||||  
Db 781 AGGTTCTGATGTGGGAAACATCTATCAGCGATTTGTCTCAGAGCATATGTCTCTGACTTAA 840  
|||||

QY 841 tttagagctctcaaatattgatgattcattgtcgtctctctgttttaccacagcttgaat 900  
|||||  
Db 841 TTTTGGAGCTCTACAATATTGATAGTCAATTTGCTGCTCTCTGTTTTACCCAGCTTCAAT 900  
|||||



9	497.4	9.4	54398	73	AC068224	Homo sapi	Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
c 10	443	8.4	72157	64	AC016449	Homo sapi	Sequence update by submitter
c 11	420.8	8.0	54398	73	AC068224	Homo sapi	On Apr 5, 1999 this sequence version replaced gi:4539617.
c 12	399	7.6	340969	81	AL512630	Mus muscu	Location/Qualifiers
c 13	232.2	4.4	2212	89	AK021757	Homo sapi	1. .5253
c 14	225.4	4.3	72157	64	AC016449	Homo sapi	/organism="Homo sapiens"
c 15	206.8	3.9	113704	92	HS267P19	Homo sapi	/db_xref="taxon:9606"
c 16	206.8	3.9	163642	79	AL353724	Homo sapi	/chromosome="13"
c 17	179.6	3.4	195380	80	AL358892	Mus muscu	/map="13q12-q13"
c 18	163	3.1	163642	79	AL353724	Homo sapi	/cell_line="LNCap"
c 19	149	2.8	150	54	G43328	WIAT-1982-S	/cell_type="cancer cells"
c 20	149	2.8	150	54	G43328	WIAT-1982-S	/tissue_type="prostate"
c 21	131	2.5	145450	73	AC068352	Homo sapi	/notes="exon 1 maps to p1 Artificial Chromosome PAC26H23; exons 2-24 map to cosmid 267p19; exons 25-34 map to p1 Artificial Chromosome PAC 49J10"
c 22	131	2.5	160962	67	AC022862	Homo sapi	1. .5253
c 23	131	2.5	167671	77	AC090081	Homo sapi	/gene="AS3"
c 24	124	2.4	69740	73	AC087748	Homo sapi	/number=1
c 25	121.4	2.3	145450	73	AC068352	Homo sapi	/gene="AS3"
c 26	121.4	2.3	162775	73	AC068352	Homo sapi	/number=2
c 27	119.2	2.3	173613	4	AC007475	Homo sapi	/gene="AS3"
c 28	119.2	2.3	192763	4	AC007475	Homo sapi	/codon_start=1
c 29	119.2	2.3	192763	4	AC007475	Homo sapi	/product="androgen-induced prostate proliferative shutoff associated protein"
c 30	119.2	2.3	192763	4	AC007475	Homo sapi	/protein_id="AAD22134.2"
c 31	115.4	2.2	69740	77	AC087748	Homo sapi	/translation="MAHSTRNDGKITPPGVKESDKISKEEMVRRLKMWVTFMD MDQDEEEKELYLNALHLASDFKHPKDVRLVACCLADIFRIYAPEYTPDK LKDIEMFITROLKLEDTKSPQNFYFYLLENIAWYKSNICFELEDNEIFQIYRT LFSVINGHGNKVMHMDLMSIICEGDTVSQELLDVNLVPAHNLNKAQYDLA KALLAKTAIEPIITNFNFQVLMGKTSISDLSEHVSLLILEYNDHSHLLSLVLPQ LEFKLSNDNEERLQVVKLLAKMFGAKDELSQNKQVCLVGRFNDIHPVIRLECY VFASHCLMNHDPDLADLTLYLKVRSHPDEEAIHVDIVISVITAAKDLILVNDHLNF VERTLDRKVRVRKEAMMGLAQIYKKYALQSAAGDAKAQIAWKDKLLHLYONSID DRLVIERFAQVMVPHNLETTERMKCLYLYATLDLNAVKALNEMWKNLHGVKID LLDLHKPQTDASVKAIKESVMYITRSLPDKGADPMKFTOVLEDEKIRKQLEVL VSPQSCQAQECVREITKNGNPKQNPPELMKFLLEIAPVHIDTESISALIKQ VNSIDGADDEGCVPTDQAIRAGLELLKVLSETHPISFSAETFEELSLACLMKDEE KVAEALQIFKNTGSKIEEDPHIRTSALLPVLHHSKSGPPQAPKALVATFIVKDL EFAQAFIEPLHKSLLDNLPHLITPLVTIGHIALLAPDQAPKALVATFIVKDL EMDRUPGKTTKLVPPDEEVSPETWKIQAIKMVMVLLGMKNHKSSTSLRLIT ILSDGDLTEOGKISKSPDMSRLRLAAGSAIVKLAQEPYHEITILEQTLQALAINDE CYOQVPAQALHKLGLSLRLPLEYMAICALCAKDPVKERRAHQCLVKNVREY LKQRAVSEKLLSLPEYVPTTHLLAHPDYVKVQDIEQLKDVKECLMFLVLEMA KNNNSHAFIRKVENIKOTDKAQGDPAKMKNEKLYTVCDVAMNINIKSKTYSLEP KDPVLPARFPTQPKNFNTKNYLPQGMKSFETPGKPTTNVLGAVNPLSAGKQSO TKSRMETVGNASSSNSSPGRIKGRLDSEMDHSENYDTMSPLPGKSDRDDS DLVRSLELKPGRKKTPTVTEREELGMDDLTKLVQEQRPKGSORSRKRGHTASESDEQ QWPEERLKEKEDILENEDEONSPKGRKGRPPKPLGGTPEEPTMTKTSKGGKSGG PPAPEEEEERQSQGTEQKSKQHRVRRRAQRAESSAESTQSTQSTPQKGRGRP SKTSPSQPKANV"
c 32	108	2.0	1489	89	AK023592	Homo sapi	174. .377
c 33	99.8	1.9	72118	10	I66494	Sequence 14	/gene="AS3"
c 34	98.6	1.9	162775	67	AC023006	Homo sapi	/number=3
c 35	87.6	1.7	167671	77	AC090081	Homo sapi	/gene="AS3"
c 36	86.6	1.6	153665	78	ALI38818	Homo sapi	/number=4
c 37	83.8	1.6	160962	67	AC022862	Homo sapi	/gene="AS3"
c 38	80.2	1.5	184535	84	CNS05TCJ	Homo sapi	/number=5
c 39	72.6	1.4	74119	71	AC036177	Homo sapi	563. 689
c 40	72	1.4	1084	53	CNS06PLB	Homo sapi	/gene="AS3"
c 41	71.2	1.4	978	53	CNS06PJK	Homo sapi	/number=6
c 42	70.8	1.3	174384	61	AC009524	Homo sapi	690. .770
c 43	70.6	1.3	66288	78	AC090825	Homo sapi	/gene="AS3"
c 44	70.6	1.3	156550	64	AC015830	Homo sapi	/number=7
c 45	70	1.3	131271	64	AC015927	Homo sapi	771. 911
RESULT 1							/gene="AS3"
HSU95825							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
MEDLINE							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
REFERENCE							
AUTHORS							
TITLE							

## ALIGNMENTS

HSU95825 5253 bp mRNA PRI 05-APR-1999  
 Human androgen-induced prostate proliferative shutoff associated protein (AS3) mRNA, complete cds.  
 U95825  
 U95825.2 GI:4559409  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 5253)  
 Geck,P., Szelei,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.  
 Androgen-induced proliferative shutoff in prostate cancer cells  
 Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)  
 2 (bases 1 to 5253)  
 Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.  
 Early gene expression during androgen-induced inhibition of proliferation of prostate cancer cells: a new suppressor candidate on chromosome 13, in the BRCA2-Rb1 locus  
 J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)  
 99229875  
 3 (bases 1 to 5253)  
 Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.  
 Direct Submission  
 Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA  
 4 (bases 1 to 5253)  
 Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.  
 Direct Submission



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 25, 2001, 14:27:17 : Search time 12186.1 Seconds  
(without alignments)  
6690.424 Million cell updates/sec

Title: US-09-512-581-1  
Perfect score: 5271  
Sequence: 1 ccgagagcccgagtgag.....gtaaaaaaaaaaaaaa 5271

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:			
1:	gb_ba1:	*	
2:	gb_ba2:	*	
3:	gb_ba3:	*	
4:	gb_in1:	*	
5:	gb_in2:	*	
6:	gb_in3:	*	
7:	gb_om:	*	
8:	gb_ov:	*	
9:	gb_pat1:	*	
10:	gb_pat2:	*	
11:	gb_ph:	*	
12:	gb_p11:	*	
13:	gb_p12:	*	
14:	gb_p13:	*	
15:	gb_p14:	*	
16:	em_ba1:	*	
17:	em_ba2:	*	
18:	em_fun:	*	
19:	em_htgo_hum:	*	
20:	em_htgo_inv:	*	
21:	em_htgo_rod:	*	
22:	em_htg_hum1:	*	
23:	em_htg_hum2:	*	
24:	em_htg_hum3:	*	
25:	em_htg_hum4:	*	
26:	em_htg_hum5:	*	
27:	em_htg_hum6:	*	
28:	em_htg_hum7:	*	
29:	em_htg_hum8:	*	
30:	em_htg_inv1:	*	
31:	em_htg_inv2:	*	
32:	em_htg_other:	*	
33:	em_htg_rod:	*	
34:	em_hum1:	*	
35:	em_hum2:	*	
36:	em_hum3:	*	
37:	em_hum4:	*	
38:	em_hum5:	*	
39:	em_hum6:	*	
40:	em_hum7:	*	
41:	em_in:	*	
42:	em_om:	*	
43:	em_or:	*	

44:	em_ov:	*	
45:	em_pat:	*	
46:	em_ph:	*	
47:	em_pl:	*	
48:	em_ro:	*	
49:	em_sts:	*	
50:	em_sy:	*	
51:	em_un:	*	
52:	em_vi:	*	
53:	gb_sts1:	*	
54:	gb_sts2:	*	
55:	gb_sts3:	*	
56:	gb_sy:	*	
57:	gb_un:	*	
58:	gb_vil:	*	
59:	gb_vl2:	*	
60:	gb_hcg1:	*	
61:	gb_hcg2:	*	
62:	gb_hcg3:	*	
63:	gb_hcg4:	*	
64:	gb_hcg5:	*	
65:	gb_hcg6:	*	
66:	gb_hcg7:	*	
67:	gb_hcg8:	*	
68:	gb_hcg9:	*	
69:	gb_hcg10:	*	
70:	gb_hcg11:	*	
71:	gb_hcg12:	*	
72:	gb_hcg13:	*	
73:	gb_hcg14:	*	
74:	gb_hcg15:	*	
75:	gb_hcg16:	*	
76:	gb_hcg17:	*	
77:	gb_hcg18:	*	
78:	gb_hcg19:	*	
79:	gb_hcg20:	*	
80:	gb_hcg21:	*	
81:	gb_hcg22:	*	
82:	gb_hcg23:	*	
83:	gb_hcg24:	*	
84:	gb_hcg25:	*	
85:	gb_prl:	*	
86:	gb_pr2:	*	
87:	gb_pr3:	*	
88:	gb_pr4:	*	
89:	gb_pr5:	*	
90:	gb_pr6:	*	
91:	gb_pr7:	*	
92:	gb_pr8:	*	
93:	gb_pr9:	*	
94:	gb_rol:	*	
95:	gb_ro2:	*	
96:	gb_in4:	*	
97:	gb_prl0:	*	
98:	em_ba3:	*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	5221	99.1	5253	97	HS095825
2	5084.4	96.5	7444	92	HS26H231
3	4906.2	93.1	5110	85	AB023196
4	1754.4	33.3	1852	97	HS050533
5	1041.4	19.8	5177	85	AB014548
6	901.6	17.1	137246	92	HS49J10
7	877.6	16.6	168487	89	AL138820
8	566.4	10.7	2079	89	AK026889
					U95825 Human andro
					AL137201 Novel hum
					AB023196 Homo sapi
					U050533 Human BRCA2
					AB014548 Homo sapi
					284572 Human DNA s
					AL138820 Human DNA
					AK026889 Homo sapi



J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1946 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 390.

## FEATURES

Location/Qualifiers  
1..391  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:856164"  
/clone\_lib="Stratagene HeLa cell s3 937216"  
/sex="female"  
/dev\_stage="HeLa S3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="vector: p Bluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3  
epithelioid carcinoma cells grown to semi-confluency  
without induction. Average insert size: 1.5 kb; Uni-ZAP XR  
vector. -5' adaptor sequence: 5' GRANTGGGACGAG 3' -3'  
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'  
BASE COUNT 112 a 81 c 64 g 133 t 1 others  
ORIGIN

## alignment\_scores:

Quality: 667.00 Length: 129  
Ratio: 5.171 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.225

## alignment\_block:

US-09-512-581-2 x AA630616/rev ..

Align seg 1/1 to reverse of: AA630616 from: 1 to: 391

```

930 LeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCy 946
|||||
389 CTTTCCCGCTTACGGCTTCCACTTGAGTATATGCGNAATCTGTGCCCTTG 340
|||||
946 sAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuV 963
|||||
339 TGCAAAAGATCCTGTAAAGGAGAGAAAGAGCTCATGTAGGCAATGTTGG 290
|||||
963 alLysAsnIleAsnValArgGluTyrLeuLysGlnHisAlaAlaVal 979
|||||
289 TGAATAATATAAATGTAAGCGGGAGTATCTGAAGCAGCATGCAGCTGTT 240
|||||
980 SerGluLysLeuLeuSerLeuProGluTyrValValProTyrThrII 996
|||||
239 AGTGAAATAATTATTGCTCTCTACAGAGATGTTGTTCCATATACAAT 190
|||||
996 eHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluG 1013
|||||
189 TCACCTTTTGGCACATGACCCAGATTATGTCAAGTACAGATATTGAAC 140
|||||
1013 lnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMet 1029
|||||
139 AACTTAAAGATGTTAAAGAAATGCTTTGTTGTTCTGGAATATTAATG 90
|||||
1030 AlaLysAsnGluAsnSerHisAlaPheIleArgLysMetValGluAs 1046
|||||
89 GCTTAAATGAAATACAGTCACGCTTTTATCAGAAAGATGCTAGAAA 40
|||||
1046 nIleLysGlnThrLysAspAlaGlnGlyProAspAsp 1058

```

|||||  
39 TATTAACAACAACAAAAGATGCCCAAGGACCAGATGAT 3



**THIS PAGE BLANK (USPTO)**